

HANDBOOK
OF
METACOGNITION
IN EDUCATION

EDITED BY

DOUGLAS J. HACKER • JOHN DUNLOSKY
ARTHUR C. GRAESSER



Handbook of Metacognition in Education

“This handbook goes a long way toward capturing the state of the science and the art of the study of metacognition. It reveals great strides in the sophistication and precision with which metacognition can be conceptualized, assessed, and developed [and] covers the gamut, including research and development on metacognition across a wide variety of subject-matter areas, as well as in more abstract issues of theory and measurement It is truly a landmark work.”

Robert J. Sternberg, From the Foreword

Providing comprehensive coverage of the theoretical bases of metacognition and its applications to educational practice, this compendium of focused and in-depth discussions from leading scholars in the field:

- represents an intersection of education, cognitive science, and technology;
- serves as a gateway to the literature for researchers and practitioners interested in one or more of the wide array of topics included; and
- sets the standard for scholarship for theoretical research and practical applications in this field.

The *Handbook of Metacognition in Education*—covering comprehension strategies, metacognitive strategies, metacomprehension, writing, science and mathematics, individual differences, self-regulated learning, technology, tutoring, and measurement—is an essential resource for researchers, faculty, students, curriculum developers, teachers, and others interested in using research and theory on metacognition to guide and inform educational practice.

Douglas J. Hacker is Professor, Department of Educational Psychology, University of Utah.

John Dunlosky is Professor, Department of Psychology, Kent State University.

Arthur C. Graesser is Professor, Department of Psychology, and Co-director, The Institute for Intelligent Systems, University of Memphis.

The Educational Psychology Series

Robert J. Sternberg and Wendy M. Williams, Series Editors

Hacker/Dunlosky/Graesser (Eds.)	<i>Handbook of Metacognition in Education</i>
Aulls/Shore	<i>Inquiry in Education, Volume I: The Conceptual Foundations for Research as a Curricular Imperative</i>
Shore/Aulls/Delacourt (Eds.)	<i>Inquiry in Education, Volume II: Overcoming Barriers to Successful Implementation</i>
Constas/Sternberg (Eds.)	<i>Translating Theory and Research into Educational Practice: Developments in Content Domains, Large Scale Reform, and Intellectual Capacity</i>
Zhang/Sternberg (Eds.)	<i>The Nature of Intellectual Styles</i>
Kyllonen/Roberts/Stankov (Eds.)	<i>Extending Intelligence: Enhancement and New Constructs</i>
Sternberg/Preiss (Eds.)	<i>Intelligence and Technology: The Impact of Tools on the Nature and Development of Human Abilities</i>
Dai/Sternberg (Eds.)	<i>Motivation, Emotion, and Cognition: Integrative Perspectives on Intellectual Functioning and Development</i>
Shavinina/Ferrari (Eds.)	<i>Beyond Knowledge: Non-Cognitive Aspects of Developing High Ability</i>
Valdés	<i>Expanding Definitions of Giftedness: The Case of Young Interpreters from Immigrant Communities</i>
Dominowski	<i>Teaching Undergraduates</i>
Corno, Cronbach, Kupermintz, Lohman, Mandinach, Porteus, Albert/ The Stanford Aptitude Seminar	<i>Remaking the Concept of Aptitude: Extending the Legacy of Richard E. Snow</i>
Ferrari (Ed.)	<i>The Pursuit of Excellence Through Education</i>
Sternberg/Zhang (Eds.)	<i>Perspectives on Cognitive, Learning, and Thinking Styles</i>
Torff/Sternberg (Eds.)	<i>Understanding and Teaching the Intuitive Mind: Student and Teacher Learning</i>
Martinez	<i>Education as the Cultivation of Intelligence</i>
Sternberg/Williams (Eds.)	<i>Intelligence, Instruction and Assessment: Theory into Practice</i>
Smith/Pourchot (Eds.)	<i>Adult Learning and Development: Perspectives from Educational Psychology</i>
Hacker/Dunlosky/Graesser (Eds.)	<i>Metacognition in Educational Theory and Practice</i>
Marton/Booth	<i>Learning and Awareness</i>

For additional information on other titles in the Educational Psychology series visit www.routledge.com/education

Handbook of Metacognition in Education

Edited by

Douglas J. Hacker
University of Utah

John Dunlosky
Kent State University

Arthur C. Graesser
University of Memphis

First published 2009
by Routledge
270 Madison Ave, New York, NY 10016

Simultaneously published in the UK
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

This edition published in the Taylor & Francis e-Library, 2009.

To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk.

© 2009 Taylor & Francis

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging in Publication Data

Handbook of metacognition in education/edited by Douglas J. Hacker, John Dunlosky, Arthur C. Graesser.

p. cm.—(The educational psychology series)

1. Cognitive learning—Handbooks, manuals, etc. 2. Metacognition—Handbooks, manuals, etc. I. Hacker, Douglas J. II. Dunlosky, John. III. Graesser, Arthur C.

LB1067.H28 2009

370.15'2—dc22

2008052651

ISBN 0-203-87642-3 Master e-book ISBN

ISBN 10: 0-8058-6353-2 (hbk)

ISBN 10: 0-8058-6354-0 (pbk)

ISBN 10: 0-203-87642-3 (ebk)

ISBN 13: 978-0-8058-6353-6 (hbk)

ISBN 13: 978-0-8058-6354-3 (pbk)

ISBN 13: 978-0-203-87642-8 (ebk)

Contents

<i>Foreword</i>	viii
ROBERT J. STERNBERG	
<i>Preface</i>	x
<i>Acknowledgement</i>	xi
1 A Growing Sense of “Agency”	1
DOUGLAS J. HACKER, JOHN DUNLOSKY, AND ARTHUR C. GRAESSER	
PART I	
Comprehension Strategies	5
2 The Role of Metacognition in Understanding and Supporting Reading Comprehension	7
MARGARET G. McKEOWN AND ISABEL L. BECK	
3 The Role of Metacognition in Teaching Reading Comprehension to Primary Students	26
JOANNA P. WILLIAMS AND J. GRANT ATKINS	
PART II	
Metacognitive Strategies	45
4 Question Generation and Anomaly Detection in Texts	47
JOSÉ OTERO	
5 Self-Explanation and Metacognition: The Dynamics of Reading	60
DANIELLE S. McNAMARA AND JOSEPH P. MAGLIANO	
PART III	
Metacomprehension	83
6 Metacognitive Monitoring During and After Reading	85
KEITH W. THIEDE, THOMAS D. GRIFFIN, JENNIFER WILEY, AND JOSHUA S. REDFORD	

7 The Importance of Knowing What You Know: A Knowledge Monitoring Framework for Studying Metacognition in Education	107
SIGMUND TOBIAS AND HOWARD T. EVERSON	
PART IV	
Writing	129
8 Metacognition and Children’s Writing	131
KAREN R. HARRIS, STEVE GRAHAM, MARY BRINDLE, AND KARIN SANDMEL	
9 Writing is Applied Metacognition	154
DOUGLAS J. HACKER, MATT C. KEENER, AND JOHN C. KIRCHER	
PART V	
Science and Mathematics	173
10 The Interplay of Scientific Inquiry and Metacognition: More than a Marriage of Convenience	175
BARBARA WHITE, JOHN FREDERIKSEN, AND ALLAN COLLINS	
11 The Enigma of Mathematical Learning Disabilities: Metacognition or STICORDI, That’s the Question	206
ANNEMIE DESOETE	
PART VI	
Individual Differences	219
12 Context Matters: Gender and Cross-Cultural Differences in Confidence	221
MARY LUNDEBERG AND LINDSEY MOHAN	
13 Teachers as Metacognitive Professionals	240
GERALD G. DUFFY, SAMUEL MILLER, SETH PARSONS, AND MICHAEL MELOTH	
PART VII	
Self-Regulated Learning	257
14 Supporting Self-Regulated Learning with Cognitive Tools	259
PHILIP H. WINNE AND JOHN C. NESBIT	
15 Effective Implementation of Metacognition	278
MICHAEL J. SERRA AND JANET METCALFE	
16 Self-Regulation: Where Metacognition and Motivation Intersect	299
BARRY J. ZIMMERMAN AND ADAM R. MOYLAN	

PART VIII	
Technology	317
17 Self-Regulated Learning with Hypermedia	319
ROGER AZEVEDO AND AMY M. WITHERSPOON	
18 Interactive Metacognition: Monitoring and Regulating a Teachable Agent	340
DANIEL L. SCHWARTZ, CATHERINE CHASE, DORIS B. CHIN, MARILY OPPEZZO, HENRY KWONG, SANDRA OKITA, GAUTAM BISWAS, ROD ROSCOE, HOGYEONG JEONG, AND JOHN WAGSTER	
PART IX	
Tutoring	359
19 Meta-Knowledge in Tutoring	361
ARTHUR C. GRAESSER, SIDNEY D’MELLO, AND NATALIE PERSON	
20 <i>In Vivo</i> Experiments on Whether Supporting Metacognition in Intelligent Tutoring Systems Yields Robust Learning	383
KEN KOEDINGER, VINCENT ALEVEN, IDO ROLL, AND RYAN BAKER	
PART X	
Measurement	413
21 Measuring Metacognitive Judgments	415
GREGORY SCHRAW	
22 Sins Committed in the Name of Ecological Validity: A Call for Representative Design in Education Science	430
JOHN DUNLOSKY, SARA BOTTIROLI, AND MARISSA HARTWIG	
<i>Index</i>	441

Foreword

Robert J. Sternberg
Tufts University

Today is April 22, 2008. It is a fairly random day—the day I happen to be writing this foreword. I mention the date, though, because it is a particularly important day for the United States. For one thing, it is the day of the Pennsylvania primary. Hillary Clinton is reported in a headline by CNN online to have said, “I have to win” the Pennsylvania primary (<http://www.cnn.com/2008/POLITICS/04/22/pa.primary/index.html>, retrieved April 22, 2008). She was right, and she won. Another headline reads “Gallup record: 69% disapprove of Bush . . . Highest negative for a president in 70 years of poll” (Gallup record: 69% disapprove of Bush, 2008). But on this same day the *New York Times* online has a smiling picture of George W. Bush with the headline, “Politicians as comics” (www.nytimes.com, retrieved April 22, 2008). On other websites, there are more smiling pictures of Bush.

It is by reading a few headlines on any random day, such as today, that one learns why metacognition is so important to psychology and to the world. Whether one likes Hillary Clinton or not, her metacognitive processes are on target: She knows that she needs to win the Pennsylvania primary, and win it big, to stay in the race for the presidential nomination of the Democratic Party. Meanwhile, George W. Bush goes through his presidency, from all appearances, oblivious to the disaster that his presidency has become in the public eye, if one views record-bottom popularity ratings as disastrous, as at least some readers might. Would you be smiling and oblivious if you learned that you were viewed as the least successful psychologist in 70 years?

Given the importance of metacognition in our lives, it is rather astonishing, as White, Frederiksen, and Collins (this volume) point out, that statewide mastery tests have all but ignored metacognition in their assessments. The skills involved may be measured indirectly, but given the kinds of advances described in this handbook, such measurement is indirect and minimal, at best. The tests are, metaphorically, shooting a .44 magnum bullet—at the wrong target.

The reason the target is wrong is that, in the long run, much of the knowledge we acquire in school that is so important in tests will be forgotten anyway. I once knew how to compute a cosecant. Today I don’t remember even what it is. I once knew what a halogen is. Those days too are long past. In my own field of psychology, I got a terrible start with a C in introductory psychology. When I sat down to write my own introductory psychology textbook, published 27 years after my ignominious grade, I discovered that most of the material covered by textbooks in 1968 was no longer even being taught in 1995. The knowledge had become largely irrelevant. The important things to acquire from the courses were not the textbook factoids, but rather, the learning to learn skills and the skills in accessing a knowledge base that form the heart of metacognition.

One could argue over who introduced the concept of metacognition. Certainly Flavell (e.g., Flavell, 1979) and Brown (e.g., Brown, 1980) would deserve much of the credit. In

the early days, metacognition was more of a curiosity and some psychologists wondered whether it was even a viable construct. Today, I think the question is not whether it is a viable construct, but rather, how it best can be understood, assessed, and developed.

This handbook goes a long way toward capturing the state of the science and the art of the study of metacognition. It reveals great strides in the sophistication and precision with which metacognition can be conceptualized, assessed, and developed. The handbook covers the gamut, including research and development on metacognition across a wide variety of subject matter areas, as well as in more abstract issues of theory and measurement. I am confident that readers will find the book as edifying and satisfying as I have. It is truly a landmark work. And most of all, I hope the leaders of our country become aware of its contents and how important metacognition is to their own leadership. What is more dangerous than leaders who don't know what they don't know? There is an answer, unfortunately. The answer is: Leaders who don't care!

Robert J. Sternberg
April 22, 2008

References

- Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. B. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453–481). Hillsdale, NJ: Erlbaum.
- Exit polls: New Dems break for Obama, late deciders for Clinton. (<http://www.cnn.com/2008/POLITICS/04/22/pa.primary/index.html>, retrieved April 22, 2008)
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, *34*, 906–911.
- Gallup record: 69% disapprove of Bush, *USA Today*, April 22, 2008 (1).
- Politicians as comics. www.nytimes.com, lead page, retrieved April 22, 2008.
- White, B., Frederiksen, J., & Collins, A. (2009). The interplay of scientific inquiry and metacognition: More than a marriage of convenience. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 175–205). New York: Routledge.

Preface

In 1998, Lawrence Erlbaum Associates published our edited volume *Metacognition in Educational Theory and Practice*. Much to our delight, interest in the topics covered in that volume has been widespread. This is reflected not only in how well this edited volume has sold, but more importantly, in how widely it has been cited in professional journals and books. In the ten years that have passed since the publication of that book, tremendous advances have occurred in both metacognitive theory and the application of theory to practice. The time is ripe to publish an updated and expanded version of that volume.

Rather than producing a second edition to our earlier volume, we decided to expand the scope of topics covered and to craft the book as a handbook rather than a textbook. As a handbook, the volume provides an opportunity for leading scholars to present focused and in-depth discussions of the role of metacognitive theory in specific areas of education. The handbook format also prominently marks our book as a major reference resource. It is a gateway to the literature for researchers and practitioners who are interested in one or more of the wide array of topics on metacognition.

Our handbook will hopefully be viewed as a standard of scholarship for theoretical research and practical application of metacognition in education. Because this handbook is targeted as a reference resource, it should appeal to a broad readership, including researchers, university professors, graduate and upper-level undergraduate students, teachers, curriculum developers, and anyone else interested in using theory to guide and inform educational practice. Our handbook could be used as the sole textbook for a graduate-level course on metacognition in general or for a more specific course on metacognition and education. It would serve equally well as a supplement to graduate courses on cognition, problem solving, learning sciences, literacy, and memory and learning.

The *Handbook of Metacognition in Education* is organized in ten sections: Comprehension Strategies, Metacognitive Strategies, Metacomprehension, Writing, Science and Mathematics, Individual Differences, Self-Regulated Learning, Technology, Tutoring, and Measurement. Each section contains two or three chapters written by leading scholars in each topic area. For example, Margaret McKeown and Isabel Beck contributed to the section on comprehension strategies; Joanna Williams and J. Grant Atkins contributed to the section on reading and writing; Barbara White, John Frederiksen, and Allan Collins contributed to the section on science and mathematics; Philip Winne and John Nesbit contributed to the chapter on self-regulated learning; and Ken Koedinger, Vincent Alevan, Ido Roll, and Ryan Baker contributed to the section on tutoring. These authors are just a small sampling of the scholars who have contributed to this handbook.

We hope that readers of this handbook will find the chapters stimulating and enjoyable. On our part, assembling the fine collection of contributors to the volume was certainly both.

Acknowledgement

We wish to thank Naomi Silverman of Routledge Publishing for her encouragement to put this volume together and for her assistance in the long process of doing so.

Douglas J. Hacker
John Dunlosky
Arthur C. Graesser

1 A Growing Sense of “Agency”

Douglas J. Hacker

University of Utah

John Dunlosky

Kent State University

Arthur C. Graesser

University of Memphis

Since the 1998 publication of our first edited volume, *Metacognition in Educational Theory and Practice*, metacognition has continued to be an important area of research, and its applications to education have continued to grow. The topics covered in our earlier volume included problem solving, reading, writing, self-regulation, technology, and meta-comprehension. In the current volume, we have expanded on those earlier topics to reflect the growing interest in the subject. We now include chapters on comprehension strategies, metacognitive strategies, mathematics and science, individual differences, self-regulated learning, technology, tutoring, and measurement.

A common theme running throughout all these chapters is the notion of agency: successful students take charge of their own learning. At a minimum, taking charge requires students to be aware of their learning, to evaluate their learning needs, to generate strategies to meet their needs, and to implement those strategies. Self-awareness, self-determination, and self-direction are the characteristics that Kluwe (1982) used when he described people as “agents of their own thinking” (p. 222). As agents of our own thinking, we construct our understanding of ourselves and the world, we control our thoughts and behaviors, and we monitor the consequences of them.

The sense of agency in metacognition puts the focus of attention clearly on the individual person. But who is this person? With the rise of cognitive science from behaviorism in the 1960s, the intent was to focus on the person as a thinking agent as opposed to an organism mechanically reacting to stimuli. Somewhat ironically, many of our colleagues in cognitive science simply replaced the empty-headed mechanics of behaviorism with a more sophisticated machine that processes information, but nonetheless follows a simple mechanistic model. With the growing interest in metacognition, we have an opportunity to look at the person in a more full-blown complexity, the self-aware agent who can construct his or her understanding of the world.

Many insights into the person (the self-aware agent) have been provided through investigations of self-concept. Much can be learned about metacognition by examining what we know about self-concept. Since the philosophical and psychological investigations of self-concept by William James (1890), self-concept has become a diverse, multifaceted construct studied in a variety of disciplines: sociology, psychology, philosophy, neuroscience, literature, and even computer science. The diversity of the construct of self-concept can be illustrated by the numerous “selves” that have been studied: minimal self, physical self, mental self, spiritual self, narrative self, extended self, ecological self, dialogical self, autobiographical self, moral self, historical self, perceived self, remembering self, remembered self, automatic self, amnesiac self, enacted self, cognitive self, and working self.

Such a diversity of selfs is daunting, but each conceptualization of self provides some insight into the self-molding and shaping agent. Consider Dennett's (1989) narrative self, which consists of an integrated collection of stories (real or imagined) that an individual tells about himself or herself to describe a specific unified agent, referred to by Dennett as the "center of narrative gravity." This narrative self provides the foundations for our predispositions, interests, and aspirations, all of which influence the kinds of activities we choose to accomplish and the amount of effort we choose to put forth in accomplishing them. Or, consider Conway's (2005) working self, which refers to an active goal hierarchy that guides and regulates pre-existing knowledge so that new knowledge can enter long-term memory. Our working self provides the bases for how we engage in online regulation of learning. Finally, consider Barresi and Juckes' (1997) dialogical self, which refers to the self that emerges as a result of interactions with individuals, groups, and culture. Our dialogical self provides input on how we perceive our social interactions and how we manage to negotiate those social interactions in the selection of goals and how we pursue them.

There is considerable overlap between the notion of agency that is derived from investigations of self-concept and the notion of agency from investigations of metacognition: comprehending the world and knowing that we comprehend, self-regulating and monitoring our thoughts, evaluating our current cognitive status in pursuit of self-imposed goals, revising our goals in light of developing cognitive and affective states, motivating ourselves, developing strategies and heuristics to make ourselves more capable of adapting to changing situations, and understanding others to gain understanding of ourselves. These are the themes that dominate the literature on self-concept and that run throughout this volume.

Each chapter describes how people guide their learning or how they can potentially be instrumental in guiding others' learning. In the opening chapters, Margaret McKeown, Isabel Beck, Joanna Williams, and J. Grant Atkins show how students understand and self-support their reading behaviors and comprehension of text. In the subsequent chapters, José Otero, Danielle McNamara, and Joseph Magliano further explore the role of the self in reading by discussing how various strategies, such as question generation and self-explanation, may support quality metacognition and effective reading. Even here, however, McNamara and Magliano argue that some aspects of metacognition are not always highly related to the use of reading strategies. They encourage others to investigate the interactive and dynamic nature of reading processes, strategy use, and strategy knowledge.

Michael Serra and Janet Metcalfe argue that students must be able to accurately evaluate their learning to take full advantage of metacognitive monitoring while studying. The importance of accurate metacognitive monitoring arises in other chapters as well, such as those offered by Sigmund Tobias, Keith Thiede, and their colleagues. For instance, Thiede, Griffin, Wiley, and Redford lament how students often have difficulties accurately evaluating how well they have learned or comprehended text materials. They then discuss factors that may limit the accuracy of judgments of text comprehension (like Serra and Metcalfe do for other metacognitive judgments) and then proceed to describe several techniques that hold promise for consistently improving students' monitoring abilities. A key assumption is that by improving students' ability to monitor their comprehension, they will be better able to discover what they do not understand and hence correct their confusion, such as by re-reading or asking a teacher or peer for assistance. Thus, enhancing this aspect of a student's metacognition may directly improve their education.

Another reason why a metacognitive approach to education has been so popular is that the basic components of metacognition can apply to almost any task that a student wants to perform. For instance, metacognitive components include (a) knowledge and beliefs about cognition, (b) monitoring cognition, and (c) regulating cognition. Knowledge and beliefs about cognition include constructs such as self-efficacy, or the degree to which a

person believes he or she can successfully complete a given task. Individual differences in self-efficacy may contribute to student successes across many domains. Moreover, beyond reading and memorizing simple concepts, students may monitor and regulate their ongoing performance on any cognitively demanding task, such as while they are composing a paper for class or when they are learning mathematics and difficult science concepts.

The pervasive nature of metacognition in student scholarship is illustrated by the range of topics covered in this handbook. For instance, several chapters highlight how metacognitive approaches have been (and can be) applied to student writing, problem solving, and learning within specific domains, such as mathematics and science. Concerning metacognition and writing, Karen Harris and her colleagues focus on how strategy instruction can benefit the quality of students' writing. The strategy instruction focuses on training several metacognitive processes, such as teaching students to monitor their performance as they write. Their instructional program, called self-regulated strategy development, has a phenomenal track record for boosting the quality of students' writing. Douglas Hacker and his colleagues further emphasize the importance of metacognition to understanding writing skills. In fact, they pose the provocative argument that defining writing as applied metacognition provides a unified framework for understanding the larger field of writing research.

Concerning science and mathematics, Barbara White and her colleagues discuss the metacognitive expertise that students will need in order to regulate their scientific inquiry as they conduct research projects. They show how training strategies and self-regulatory activities can boost students' learning about how to explore and inquire about science. This program appears especially useful for lower-achieving students. Annemie Desoete focuses on lower-achieving students in mathematics. Students with mathematical learning disabilities appear to have relatively normal intelligence, yet they demonstrate a specific deficit in mathematics. Desoete entertains whether metacognitive problems contribute to these disabilities, and how other aids can benefit these students when metacognitive interventions fail. Both chapters underscore the value of a metacognitive approach for better understanding, and potentially remediating, the difficulties that many students encounter as they tackle difficult problems in science and mathematics.

Students do not always study and complete homework in a solitary environment, but rather their learning is supported by other students, teachers, and tutoring systems. In the present handbook, multiple chapters focus on how other people and technologies can help promote student learning and metacognition. Phil Winne and John Nesbit describe multiple ways that students' metacognition can go awry as they study. Students often inaccurately monitor, they are unduly influenced by fallible heuristics, and they may not seek help in a useful way. Most relevant here, these authors describe how computer technologies can be used to promote successful metacognition and learning, such as by prompting students to deeply process to-be-learned materials and to assess their comprehension while on task. Similarly, Ken Koedinger and his colleagues offer a detailed discussion of their research with computer-based tutoring systems. Their systems provide several kinds of metacognitive support that have been demonstrated to promote learning that is retained over the long term, and that is evident on transfer tests. Roger Azevedo and Amy Witherspoon synthesize research on hypermedia learning with an information processing theory of self-regulated learning. In doing so, they offer guidelines for how to support effective self-regulated learning with hypermedia.

Although computer-based tutors and other computer technologies are being used to promote effective metacognition and student learning, what is striking is how little is known about the metacognitive abilities of others who have driven education practices for centuries: teachers and tutors themselves. In groundbreaking chapters, Art Graesser

and his colleagues consider the metacognitive abilities of tutors, and Gerald Duffy and his associates discuss the metacognition of teachers. Both areas are largely unexplored. For instance, as noted by Duffy et al., “While researchers and educators claim frequently that teachers are metacognitive, detailed characterizations based on empirical qualitative or quantitative evidence are scarce” (p. 553). In both chapters, the authors do an excellent job of not only discussing key metacognitive issues in each area (e.g., what metacognitive illusions tutors may hold that can undermine their effectiveness), but they do so in a manner that will stimulate systematic research programs in these relatively unexplored areas.

In this handbook, you also will find a number of chapters that are meant to offer foundations in conducting metacognitive research. Sigmund Tobias and Howard Everson describe the knowledge monitoring assessment tool. Using this tool, Tobias and his colleagues have shown that accurate knowledge monitoring, namely being able to distinguish between what one does versus does not know, is related to educational outcomes across a variety of domains. As important, Gregory Schraw provides an overview of basic measurement issues in metacognitive research that is relevant to understanding the measures described in other chapters in this handbook. Finally, John Dunlosky and colleagues argue that the generalization of metacognitive research, and research in general, to real-world contexts might be better conceptualized and understood from the framework of representative design rather than ecological validity.

Our brief overview in this chapter merely scratches the surface of the number of issues and topics raised about metacognition in this handbook. Other topics include analyses of metacognition and individual differences in gender, culture, and motivation, while other contributors consider the potential role of metacognition in teaching students to read and even the educational benefits of having students teach a computer agent. Despite the distinct topics that are discussed within any given chapter, most chapters herald a common theme: people’s metacognitive processes, when appropriately engaged, can be used to scaffold effective learning, problem solving, and comprehension. Our self-concepts motivate us toward successes or failures. With continued efforts to understand how to promote effective metacognition and healthy self-concepts, research on metacognition in education promises to further improve student education. Perhaps most important, each chapter in this handbook highlights how these promises are currently being actualized in the laboratory and in the classroom.

Based on the growth and impact of metacognitive research in education since 1998, we expect that many new advances will arise in the next decade. These advances will further highlight both the promises and limits of using metacognition to improve student education. We hope that the next generation of researchers will continue to engage in active discussion and research that explores the utility of metacognition in education. Their insights will no doubt inspire publication of the next handbook on metacognition in education.

References

- Barresi, J. & Juckes, T. J. (1997). Personology and the narrative interpretation of lives. *Journal of Personality*, 65, 693–719.
- Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, 53, 594–628.
- Dennett, D. C. (1989). The origins of selves. *Cogito*, 3, 163–173.
- James, W. (1890). *The principles of psychology*. New York: H. Holt.
- Kluwe, R. H. (1982). Cognitive knowledge and executive control: Metacognition. In D. R. Griffin (Ed.), *Animal mind—human mind* (pp. 201–224). New York: Springer-Verlag.

Part I

Comprehension Strategies

2 The Role of Metacognition in Understanding and Supporting Reading Comprehension *

*Margaret G. McKeown and
Isabel L. Beck*
University of Pittsburgh

What do we want students to do as they read? Do we want them to be consciously aware of their processes in order to have control over them? What is the level of awareness and control that best supports comprehension? We will move toward addressing these questions, beginning by reviewing the concept of metacognition through its historical roots, as described by Brown, Bransford, Ferrera, and Campione (1983), and discussing how metacognition has found application to reading in the form of strategies instruction. We then describe an alternative approach to comprehension instruction and discuss results of a study that compared the two approaches.

Origins of Metacognition and its Route to Reading

Brown et al. (1983) view the concept of metacognition as having four historical roots, each of which has provided foundation for approaches to strategies instruction, which we will take up in the next section. The first root is the issue of verbal reports as data—how reliable are people’s reports of their thinking processes? What can we express about what we know, or how does what we can express relate to what we know? The second root is the notion of executive control, which is derived from information processing models. These models feature a central processor that can control its own operations, which include planning, evaluating, monitoring, and revising. The third root is self-regulation, processes by which active learners direct and continuously fine-tune their actions. The fourth root that Brown et al. see underlying metacognition is what they call other regulation, or the transfer of control from other to self. This kind of regulation is based on Vygotsky’s theory that all psychological processes begin as social and are then transformed through supportive experience to the intrapersonal.

A number of the components of metacognition that Brown et al. discuss within the four roots have relevance for reading. Actions such as self-regulating, planning, evaluating, and monitoring align well with what researchers have come to see as the processes in which readers need to engage in order to achieve successful comprehension. As Baker and Brown (1984) put it: “Since effective readers must have some awareness and control of the cognitive activities they engage in as they read, most characterizations of reading include skills and activities that involve metacognition” (p. 354).

Such characterizations of reading developed as a cognitive processing perspective on reading took hold. Initially, information-processing models were applied to fundamental

* We gratefully acknowledge the Institute for Education Sciences of the U.S. Department of Education for its support of the research described in this article. The opinions expressed do not necessarily reflect those of the institute, and no official endorsement should be inferred.

psychological processes of thinking and learning. Studies of reading from this perspective highlighted reading as a complex mental process with various interacting subprocesses. With this perspective, it became clear that precisely because of the complex nature of reading, many opportunities existed to intervene and help students develop more effective processes.

A need for new approaches to intervene in reading development became apparent from an abundance of research in which students' performance on various post-reading tasks was seen to be inadequate. More specifically, findings showed that students had difficulty with an array of reading tasks, including ones as fundamental as identifying main ideas, at least in stories beyond simple folktales (Brown & Smiley, 1977; Smiley, Oakley, Worthen, Campione, & Brown, 1977). Aspects of developing summaries challenged students in grades 5, 7, and 10, although older students did well on simpler tasks such as deleting information. But students at all levels had trouble when they needed to put text information in their own words or draw inferences (Brown & Day, 1983). Identifying inconsistencies or confusions that had been planted in texts proved difficult for students from elementary (Markman, 1979) and middle school (Garner, 1981; Garner & Kraus, 1981–1982) to college age (Baker, 1979). Other aspects of monitoring comprehension were also absent from readers' performances, especially for poor readers (Brown, Campione, & Barclay, 1979) or when text was more difficult (Olshavsky, 1976–1977).

Overall, this body of research indicated that many students do not carry out the processes of reading with much awareness of the actions needed to meet comprehension goals (Brown, Armbruster, & Baker, 1986; Wagoner, 1983). That is, students often finished reading a text not really knowing if they had understood it and not able to do anything about it.

Our own research was among the body of work that examined students' reading comprehension and often found it lacking (Beck, Omanson, & McKeown, 1982; Omanson, Beck, Voss, & McKeown, 1984; McKeown & Beck, 1994). Our research comprised a series of studies with students from second to eighth grades on both narrative and social studies texts. In studies of second- and third-graders' reading narratives from their basal reading books, findings suggested that the students encountered difficulties in comprehension when using the text and lessons as presented in instructional materials. We found that students did significantly better when they were given greater support than provided in the reading program lessons for during-reading questions (Beck et al., 1982).

Our studies of students reading social studies texts focused on a sequence of events leading to the American Revolution that is typically studied in fifth and eighth grades. The research with fifth-graders presented four short sequential passages from the students' social studies book and asked students for recall after each text passage. The passages covered the French and Indian War, no taxation without representation, the Boston Tea Party, and the Intolerable Acts.

Our studies were initiated at the point where students would have been reading the pre-Revolutionary text sequence in social studies class—so they were seemingly prepared to encounter the information. However, we found many rather startling examples of lack of understanding that suggested that students were not attending to what they read in thoughtful, reflective ways. For instance, in recalling the text about the French and Indian War, one student said that the issue was that the Indians wanted to be free from the British, and “they wanted to start their own land,” and that after the French and Indian War and the American Revolution “the Indians won it and they got their freedom” (McKeown & Beck, 1994). In discussing the issue of no taxation without representation, students frequently omitted the major ideas that Britain levied taxes on the colonies to pay for the French and Indian War, and that the colonists did not want to pay any taxes levied by

Britain. For example, after reading the approximately 200-word description of the events around this issue, one student recalled only that it was about Britain's Parliament passing laws and that after a few years things quieted down. Another student read the word *colonists* as *colonel* and created a scenario based on her fictitious military officer, explaining that "the colonel was making a law and the British soldiers didn't like it" but that they could not end the laws because "this colonel was too strong for them."

The foregoing examples are illustrative of findings from studies of students' reading, including others in the literature as well as our own. Students seemed often unaware that what they were getting from text did not make sense, as demonstrated in our "colonel" example. In fact, given that students often "recalled" information that was so much at odds with the text, one wonders whether students understood that making sense of what they read was the goal of the enterprise. Metacognitive aspects of the process, such as evaluating, monitoring, and revising what was taken from a text seemed to match with what was missing from students' reading. So a major focus in reading research became how to help students acquire metacognitive abilities. This was implemented for the most part by turning to the components of metacognitive ability and directly teaching students what those were and how to engage them in interactions with text, that is, strategies instruction. Strategies were conceived as representing routines that successful readers engaged in as they read in order to keep their comprehension on track.

Strategies Literature

A large and sprawling literature on strategies has developed, including several programs of strategies instruction framed around deliberate sets of strategies, studies on combinations of various strategies, and studies on single strategies. An extensive review of this literature was part of the National Reading Panel (2000) report. The report's conclusion reflected what is generally regarded as consensus in the field, that is, instruction in reading strategies has been highly successful overall. The report points to the use and coordination of multiple strategies as particularly effective "where teachers and readers interact over text" (pp. 4–46). Just what success of strategies instruction means will be explored in a later section. In this section we provide an overview of the strategies literature.

In general, the literature has moved from single strategies studies to use of multiple strategies (National Reading Panel, 2000), however, programs built around sets of strategies designed by Palincsar and Brown (1984), Paris, Cross, and Lipson (1984), Duffy and Roehler (1989), and Pressley et al. (1992) were initiated early on. Each of these programs can trace its foundation to the historical roots of metacognition described by Brown et al. (1983).

Palincsar and Brown (1984) approached the development of instruction for reading comprehension through the metacognitive root of what Brown et al. (1983) characterized as "other regulation," learning that is initiated by social interaction with a more expert other and gradually transformed into intrapersonal ability. In Brown et al. (1983), the authors describe a dialogue procedure that seems to be a precursor to Reciprocal Teaching (RT) in which a tutor and tutee share a text, switching roles to summarize, clarify, predict, and question each section. As RT was further developed, it was implemented in classrooms with the teacher and students taking turns to perform the four strategies—summarizing, clarifying, predicting, and asking a question—after each segment of text was read. The introduction of RT was a highly significant development in work on reading comprehension. Palincsar and Brown's work marked a shift in the role of the teacher from evaluating comprehension to guiding the process and a shift of the work of comprehension instruction from a separate activity focused on worksheets to activities embedded in real reading

(Rosenshine & Meister, 1994). The four strategies identified for RT set the stage for a body of work on strategies instruction.

Pressley and his colleagues (Symons, Snyder, Cariglia-Bull, & Pressley, 1989) have approached the issue of assisting students with comprehension through the metacognitive root that Brown et al. (1983) label executive control. Pressley et al.'s approach was based on models of thinking (Baron, 1985; Sternberg, 1979, 1982) that are derived from information processing conceptions of competent performance. According to these models, competent thinking is a series of search processes executed with stored knowledge and schema toward solving a problem. Within these processes, thinkers employ strategies such as identifying their goal, monitoring their progress, and evaluating evidence.

Considering these models, the reasoning was that providing young students with some procedures they could employ while reading could facilitate their comprehension. This line of thinking led Pressley and his colleagues to develop Transactional Strategies Instruction (TSI). Compared to RT, TSI has a larger set of strategies at its core, including predicting, altering expectations as text unfolds, generating questions and interpretations, visualizing text ideas, summarizing, and attending selectively to important information. TSI proceeds with the teacher first explaining new strategies, modeling use by explaining their thinking processes, and explaining why and when to use the strategies. As students practice the use of strategies during reading, the teacher provides feedback and cues students to transfer the strategies to other situations. This kind of cycle takes place over time, with abundant practice provided on a few strategies before additional ones are introduced (Pressley et al., 1992).

Paris and his colleagues focus their instructional approach, Informed Strategies for Learning (ISL), on developing awareness of the goals of reading and the value of using strategies to pursue those goals. They discuss two aspects of metacognition, the first being knowledge of the role of strategies and the second being an executive function that orchestrates higher-order cognitive processes. These aspects seem to reflect two roots that Brown et al. (1983) discuss, the issue of verbal reports as data—or what do we know about our mental processes—and the executive aspect. Paris et al. (1984) contrast their approach with those based on other regulation, saying that in those cases students are taught to adopt roles and cognitive activities of the teacher. Their instruction is designed to teach students to evaluate, plan, and regulate as they build awareness of their processing. The strategies taught in ISL include understanding the purposes of reading, activating background knowledge, allocating attention to main ideas, critical evaluation, monitoring comprehension, and drawing inferences, as well as others that are not enumerated. As strategies were introduced, students were stimulated to think about them and reflect on the goals of reading and the actions they could take to reach the goals. “Questions, dialogues, analogies, and modeling were all used to stimulate awareness” (p. 1243). Paris et al.'s ISL is similar to Pressley et al.'s TSI in the focus on executive functions of coordinating when and where to use strategies and in some of the activities engaged. But Paris's ISL seems to have a stronger emphasis on explicit knowledge of strategies, building awareness of the reading process, and the appropriateness of applying strategies.

The work of Duffy and Roehler and their colleagues (1987) also strongly emphasizes the explicit strategies knowledge component, hypothesizing that greater metacognitive awareness will eventually yield higher reading achievement. Duffy et al. emphasize the metacognitive root of self-regulation, discussing the importance of self-monitoring and that other approaches do not accomplish it. In their instruction, Duffy et al. focus on strategies to remove blockages to comprehension. They emphasize the role of direct explanation of strategies and their importance and of explicit modeling in the instruction. They think of the modeling as a way to teach reasoning, which is their goal in presenting

cognitive strategy instruction. Modeling is intended to make thinking processes visible, thereby giving students substantive information about how to be strategic. The language of instruction in this approach (Duffy & Roehler, 1989) is specific, direct, and elaborate.

Beyond these programs that have had sustaining impact on the instructional strategies field, there are additional instructional strategies programs (see, for example, Anderson & Roit, 1993; Block, 1993; Klingner, Vaughn, & Schumm, 1998) as well as numerous smaller studies on a variety of strategies—implemented both individually and in combinations. The National Reading Panel (NRP) report, in addition to providing a picture of overall success with strategies instruction, identified seven individual strategies that the panel found to be “effective and most promising for classroom instruction” (pp. 4–42). These were: comprehension monitoring, cooperative learning, graphic and semantic organizers, question answering, question generation, story structure, and summarization.

Strategy Instruction: Issues and Caveats

Although strategy instruction has been labeled a success, there are many issues remaining. For example, what is a strategy? Based on its metacognitive heritage, it seems that a strategy could be described as a routine that represents a specific mental processing action that is part of a larger, complex process executed toward a goal—such as understanding what one has read. But several of the successful strategies identified in the NRP report would not seem to fit that description. For example, is using graphic organizers a strategy? It seems like a technique that can be used for studying, but it is not a processing action. Can cooperative learning plausibly be called a strategy? It would seem to be more accurately described as a structured format for classroom learning. And comprehension monitoring seems more complex than a single strategy. Given this lack of clarity in determining what a strategy is, what does it mean to say that strategy instruction works? What is it that is working?

Even if we accept as effective strategies the categories that NRP presents, the issue of defining a strategy remains, although at a different level. That is, the instructional activities within a strategy category are not consistent from study to study. In some cases, the same strategy label is given to quite different activities. Consider the variety of tasks and activities included in the studies of summarizing, for example. In one study, students are taught steps for creating a summary, including selecting main information, deleting trivial information, and relating to supporting information (Rinehart, Stahl, & Erickson, 1986). In another summarization study, students are taught text structure categories (Armbruster, Anderson, & Ostertag, 1987), and in another to use “who?” and “what happened?” questions for each paragraph of text (Jenkins, Heliotis, Stein, & Haynes, 1987).

A similarly inconsistent picture emerges from examination of the studies labeled comprehension-monitoring. A study in that category by Schmitt (1988) instructs students to activate prior knowledge, set purposes, generate and answer prequestions, form hypotheses, verify or reject hypotheses, evaluate predictions, and summarize. In contrast, a study by Miller (1985) in the same category included teaching students to ask themselves questions as they read, such as “is there anything wrong with the story?” and to underline problems they found. Thus, not only do activities under comprehension-monitoring vary widely, but studies on this strategy also include activities that are the domain of other strategy categories, such as summarizing and asking questions. The result is that the sum total of studies leaves us unable to discern what makes a particular strategy effective, or what activities instantiate a particular strategy.

Another issue related to understanding how strategy instruction functions is that examples of students’ interactions with text, beyond a line or two, are rarely included in

reports of the research. In studies of strategy instruction, information is typically provided about the initial instruction of the strategy, and some examples of modeling and initial student practice are presented. But we see few examples of the use of strategies playing out as teachers and students interact with texts. It seems, then, that the work on strategies has left us with fundamental questions of why strategies work and what the necessary parts of strategy instruction are that bring effects.

The issue of what it is about strategy instruction that makes it effective seems to be reflected in comments that a number of reading scholars have made. The comments seem to go to the heart of seeking what is essential for comprehension instruction. For example, Carver (1987) has suggested that the positive effects of strategies may be due to increased time spent reading and thinking about text rather than to the specifics of that instruction. Pearson and Fielding (1991) pondered whether there would be a need for instruction in strategies if student attention could simply be focused on understanding text content. Seeming to address this issue, Gersten, Fuchs, Williams, and Baker (2001) suggest moving from explicit strategies toward more fluid approaches to comprehension development centered on getting students to read in a more thoughtful way. In a similar vein, Sinatra, Brown, and Reynolds (2002) raise the question of whether it is more effective to explicitly teach comprehension strategies or to teach students to approach reading with a problem solving perspective. Sinatra et al. see an advantage in the latter, expressing a concern that focusing on implementing strategies may draw students' attention away from comprehension.

Sinatra et al.'s concern has been similarly expressed by others who question whether aspects of strategy instruction might be counterproductive. For example, Winograd and Johnston (1987) questioned whether direct instruction on strategies might give young readers too many options to make decisions about during reading, rather than encouraging understanding. Pearson and Dole (1987) cautioned about the possibility of asking students to pay so much attention to their use of strategies that it "may turn relatively simple and intuitive tasks into introspective nightmares" (p. 162).

Developing Another Approach to Comprehension Instruction

In our own research, when findings led us to conclude that students needed to engage with what they were reading and be aware of whether the ideas were making sense to them, we initially considered teaching students to use strategies. But as we tried to specify how we might do that, we became concerned that focusing on strategies might turn students' attention to the specific procedures for implementing the strategies rather than directly to the text content. So in trying to develop a way to work with students, we first asked students to read a text, sentence by sentence, and think aloud about each sentence. As they responded, we probed for more information. We began to see a pattern of probes that seemed most productive: when we asked what the author was trying to say, or what the author meant; when we asked simply "what was that all about?" We noticed that in response to such probes, students tended to explain what they had understood and how they were trying to make sense of what they read.

We considered the kind of responses we were getting from students in light of theoretical descriptions of the reading process from the work of cognitive scientists such as Kintsch and van den Broek (Kintsch, 1988; Kintsch & van Dijk, 1978; van den Broek, 1990). Within this cognitive processing perspective, reading proceeds by selecting and focusing on important content in the sentence currently being read, carrying that information forward, and figuring out how to connect it to content subsequently read. When satisfaction with the level of comprehension is reached, reading proceeds to the next sentence and the

process repeats, toward creating a coherent representation of the entire text. Coherence depends on the connections made along the way, chiefly causal and referential relations.

Working from this cognitive processing perspective, we developed our approach to comprehension, Questioning the Author (QtA) (Beck & McKeown, 2006; McKeown, Beck, & Worthy, 1993). In QtA, text is read and the teacher stops at planned points to pose open questions such as “What is the author trying to say?”, “What’s happening now?”, “How does this connect to what we read before?” These kinds of questions invite students to consider the important content and build connections among the key ideas in that content toward developing a coherent representation. The open questions initiate talk about the text, but the grist of comprehension development really comes from the interactions that follow, as teachers take their cues from students’ responses and encourage students to build on those ideas, with probes such as “That’s what the author said, but what does the author mean?”, “You told us . . . what’s that all about?”, and “How does that fit in with what the author already told us?” A QtA discussion essentially mimics a successful comprehension process by helping students to select important content and connect it as reading proceeds.

Research on classrooms implementing QtA has found significant changes in the kind of talk about text and improvements in comprehension and comprehension-monitoring (Beck and McKeown, 1998; Beck, McKeown, Sandora, Kucan, & Worthy, 1996; McKeown, Beck, & Sandora, 1996; Sandora, Beck, & McKeown, 1999). The talk surrounding text being read changed in that both teachers’ questions and student responses shifted from a focus on retrieving information to seeking and building meaning. Teachers’ responses to students changed from evaluating or repeating the response to responding in ways that extended the discussion, and the proportion of student talk greatly increased (Beck et al., 1996; McKeown et al., 1996). The Beck et al. and Beck and McKeown studies included an individually-administered comprehension task on a novel text passage which measured growth in comprehension-monitoring and comprehension of the text. Both the Beck et al. (1996) and Beck and McKeown (1998) studies showed advantages for QtA students; in the first, students improved in monitoring, and in the second, both monitoring and comprehension increased. The Sandora et al. study (1999) compared the effects on students’ comprehension of QtA and Junior Great Books, and found that students in the QtA condition had greater recall and higher scores on answers to interpretive questions than those in the Junior Great Books group.

Comparing Approaches to Comprehension Instruction

Strategies instruction and the QtA approaches have common features as well as distinctions. Both approaches try to engender deep engagement with reading. Both approaches certainly intend that students understand the content of a text with which they are working. The distinction is that a strategy approach encourages students to think about their mental processes and on that basis to execute specific strategies with which to interact with text. In contrast, QtA attempts to engage students in the process of attending to text ideas and building a mental representation of the ideas, with no direction to consider specific mental processes. At issue is whether understanding will best come about by applying explicit strategies or whether it comes from directly examining and working through the content.

Overview of a Comparative Study

In a recent study, we compared the effects of strategy instruction to QtA (McKeown, Beck, & Blake, in press). To conduct the study, we developed sets of standardized lessons for

strategies and QtA around a common set of texts for fifth grade. The study ran for two consecutive years. In the first year the lesson materials were based on five narratives from the basal reader in use in the school district. In the second year, these same story lessons were used again, and we added three expository texts. For this chapter we will focus on the second year of the study, because methods and results are very similar across the two years except that the second year had additional texts and measures.

The study included all 119 fifth-graders from one school in a low-performing urban district. This involved six classrooms and their teachers, two classrooms in which teachers taught strategies lessons, two classrooms in which teachers taught QtA lessons, and two classrooms in which lessons using the basal reader material were taught, serving as our comparison group. In this chapter we will confine our discussion to the results from the strategies and QtA classrooms.

To develop the lessons for the texts we had selected, we first needed to define what our strategy instruction would consist of. We began by considering which strategies had been highlighted in two major recent reports on reading, the National Research Council's (NRC) *Preventing Reading Difficulties in Young Children* (Snow, Burns, & Griffin, 1998) and the NRP (2000) report discussed earlier. The NRC report (Snow et al., 1998) focuses on summarizing, predicting events and outcomes of upcoming text, drawing inferences, and monitoring for coherence. The NRP (2000) report lists comprehension monitoring, summarization, question-generation, question-answering, cooperative learning, graphic and semantic organizers, and multiple-strategy teaching. Next we considered which of those procedures might be most naturally called on as a reader works through a text to understand the content. Our thinking was that readers tend to summarize important information as they move through text, they develop a sense of what may be coming next, they need to draw inferences to create connections, and they may well form questions to check that they are on track. Additionally, effective readers monitor their understanding and take steps to remedy the situation if they do not understand. We thus selected summarizing, predicting, drawing inferences, question-generation, and comprehension monitoring as the strategies for our lessons. We developed the strategies instruction with input from strategies experts in the field.

For lessons in all three conditions we followed a similar format, which we scripted for the teachers. We chose stopping places in the text, which were very similar across the approaches, and developed questions for the teacher to pose (in the case of QtA and the comparison) and procedures to prompt students to implement a specific strategy for the strategies condition. The scripts also included suggestions on following up student responses, in case students did not address key information in their initial responses.

A stop in a Strategies lesson, for example, might begin with the teacher saying "This is a good place to stop and summarize." After a student responds, follow-up prompts suggested for the teacher include: "Was that a good summary?", to have other students evaluate and add or revise, and "What do we do when we summarize?", to have students review the thinking that goes into summarizing.

A stop in a QtA lesson might ask "What just happened?" with a follow-up provided in case key information was not elicited, such as, "Why might that be important?"

The lessons were presented over nine weeks. For the first five weeks, a narrative lesson was presented at the beginning of each week in all six classrooms. Two of the narratives were completed within one classroom period of 45 minutes; the other three were divided into two parts, which were delivered on consecutive days. Three additional expository lessons of about 45 minutes each were implemented in the classrooms after the completion of the five narrative lessons.

Measures and Outcomes

We used a variety of measures to assess the outcomes of the lesson conditions. This included a comprehension-monitoring task, a comprehension test for each story based on Royer's sentence verification technique (Royer, Hastings, & Hook, 1979), recall of texts used in the lessons, and recall of a transfer text. Our analyses showed no differences between groups on the comprehension-monitoring task or the sentence verification task. Differences were found in recall of both lesson and transfer texts in favor of the QtA group. We view the recall differences as important, because recall is a productive measure that is usually considered to capture a higher level of comprehension than a test that requires recognition, such as sentence verification. For the present discussion we will focus on the recall tasks because the recalls reveal some qualities of students' comprehension that are key to comparing the two approaches.

Recall of Lesson Texts

We collected individual oral recalls for two of the five narrative texts and two of the three expository texts that were used in the classroom lessons. After lessons for each of the texts were completed, we met with each student and asked them to "tell us the whole story of [title]." Recalls were recorded, then transcribed and scored. A scoring scheme for the recalls was developed to capture quality of recall based on weighting text units by level of importance—major, supportive, and detail—and awarding students a score based on which level and how many units they recalled. The units and weighting were determined by breaking the text into approximately clausal units and analyzing the units into major, supportive, and detail, based on a procedure developed by Omanson (1982).

We found differences for three of the four text recalls—for both of the narrative texts and one of the expository texts. The differences indicated moderate effect sizes: for one of the narratives, $d = 0.660$, for the other narrative, $d = 0.481$, and for the expository text, $d = 0.509$.

Recall of Transfer Text

The ultimate goal of approaches to comprehension instruction that aim to affect the way students think about text as they read, as do strategy instruction and QtA, is to affect students' thinking when they are reading independently. Thus in our study we designed a transfer task to explore how well students under each approach were able to do on their own. The task involved a sequence of five text lessons that provided gradually less support from the teacher and discussion. The lessons were introduced with the teacher reminding students to do the same kind of thinking on their own as they had been doing during class discussions. Thus, in QtA classrooms, the teacher reminded students that they were trying to understand what they had just read and prompted them to describe what they were thinking about, while in the Strategies classrooms, students were asked to choose a strategy that would fit best at each stopping point, followed by discussion of how the strategy had been applied.

As the lessons progressed, the teacher began to provide prompts that were followed not by discussion but by giving students the opportunity to think silently about how they would respond. The fourth text included only prompts for thinking, as did a fifth text that was then assessed for oral recall.

The assessed text was a 973-word hybrid text, that is, a text that combined expository

and narrative genres. The text was about the growth of factories during the Industrial Revolution with an embedded narrative about girls who went on strike over working conditions and wages. Analysis of the student recalls showed that students in the QtA classrooms had higher comprehension scores and the difference again indicated a moderate effect size, $d = 0.490$.

A Closer Look at Students' Comprehension

Are the differences in recall between the Strategies and QtA groups meaningful? We address this question first by examining the recalls themselves. We then explore the transcripts of the lessons to see if the roots of the differences can be identified in how the lessons played out.

For our examination of the recalls, we first present the transfer text recalls and then the lesson text recalls. To provide a systematic examination, we developed prototypical recalls for each text, which represent what a typical student recall would look like for the strategy group and the QtA group. The prototypical recalls are based on the most frequently recalled text units. The number of units in each group's prototypical recall was determined by the average number of units recalled by that group.

Prototypical Recalls of Transfer Text

Before considering students' recalls, we present a summary of the transfer text:

In the middle of the 1800s work that was once done by people in their homes was now done in factories. Small towns changed into large cities. Lowell, Massachusetts, was one such town, which soon had many textile factories. The factory owners hired children to work in the factories, because they could pay them lower wages. These children, mostly girls, worked to make money for their families. They worked long hours under poor working conditions. Harriet Hanson was one of these working girls. As more factories were built and competition increased, the factory owners lowered the wages and made the girls work faster. The girls started to talk about striking. As the girls began to organize to strike, Harriet was interested but worried about disgracing her family if she lost her job. The girls thought striking would force the owners to pay them more. When the appointed day arrived, the girls left the factory. Harriet was proud to be among them.

The prototypical recalls of the Strategy group and the QtA group are below. The text in bold represents units that are unique to that group's recall. Notice that the recalls of both groups represent a competent understanding of the text. That is, they capture the major outlines of the content and do not include trivial details. The two groups have most of the recalled units in common. But the uniquely recalled units of each group play different roles in the text, which we will discuss below.

Strategies group prototypical recall:

The mill owners hired children, mostly girls, to work. The girls worked in these mills to get money for their families. Harriet was one of the girls that worked in these factories. Soon, more factories started to pop up throughout New England. The factory owners cut the pay of the workers **and they made them work harder**. The girls were upset. **They thought about striking**. Harriet was nervous about striking. However, Harriet and the other girls decided to strike.

QtA group prototypical recall:

There were factories. The mill owners hired children, mostly girls, to work. These girls worked in the mills to get money for their families. Harriet was one of the girls that worked in these factories. **She woke up early to go to work.** Soon, more factories started to pop up throughout New England. **The town mills started to have competition.** The factory owners cut the pay of the workers. The girls were upset. **They thought that the conditions would improve if they went on strike.** Harriet was nervous about striking. **She didn't want to disappoint her family by losing her job.** However, Harriet and the other girls decided to strike **and they left the factory.**

The Strategies group recall has two unique units. One of them, “They thought about striking,” is implied in a unit that the QtA group has recalled, “They thought that the conditions would improve if they went on strike.” So it does not represent any additional recalled content. The other unit, “and they made them work harder” is an elaboration of the reasons for striking.

Now consider the units that the QtA group uniquely recalled and the role of that information in the text. Four of the units supply links in a causal chain: “The town mills started to have competition” represents what causes the factory owners to lower the workers’ pay, which in turn motivates the girls to strike. The unit “She didn’t want to disappoint her family by losing her job” represents the cause of Harriet’s hesitation about striking. In the unit “They thought that the conditions would improve if they went on strike” the information that is uniquely recalled, about the potential for conditions to improve, is the girls’ goal in striking. The unit “and they left the factory” is the concluding event that represents the girls’ accomplishing their plan of striking. Another unique unit provides the starting premise for the chain of events that unfolds: “There were factories.” Only the unit “She woke up early to go to work” has no role in the causal chain.

So of the six units that the QtA students uniquely recalled, five connect causes with actions, causes with consequences, or actions with motivations. These are the kinds of relationships that are at the core of a successful comprehension process. The recall suggests that experience with QtA may prompt students to seek connections between ideas and events as they read.

The text just discussed is the one that students dealt with on their own, with no accompanying discussion and follow-up prompting by the teacher, as occurred in the regular lessons within each condition. But examining the recalls that the students gave in response to the lesson texts suggest that the seeming predilection of QtA students to produce causally connected recalls was developed in the lessons.

Prototypical Recalls of Lesson Texts

The recalls for the three lesson texts for which there was a difference between the Strategies and QtA students show qualitative differences in the units most frequently recalled that are similar to the recalls of the transfer texts. Here we will highlight those differences by discussing the differences in the prototypical recalls for each condition. We begin one of the narratives, *Off and Running* (Soto, 2005). Below we present a summary of the story:

A girl, Miata, is running for fifth grade president against a boy, Rudy. Miata’s campaign promises include cleaning up graffiti on school grounds and planting flowers, while Rudy promises extra ice cream days. Miata fears she will lose the election and

seeks someone famous to endorse her campaign. Her mother tells her of a relative who was mayor of a Mexican town, and Miata goes to visit the woman, Dona Carmen. The woman tells her about her own campaign goals to better the education in her town. Miata begins to see that she is offering something worthwhile. Dona Carmen offers to help her by giving her flowers for the school, and the story ends with Miata picturing how the school will look if she gets to fulfill her own promises.

The recalls for the strategy and QtA groups show that, as with the transfer text, both groups produce typical recalls that include the major outlines of the story. This includes that Miata is running for class president against Rudy, what the two promise in their campaign, that Miata goes to visit the former mayor to ask her for election advice, and that the mayor offers her help. The strategies group recalls typically include no units that were not also included in the QtA recalls.

Unique to the QtA recalls, however, are eight units. These include:

- that Miata fears she is losing the election, which is the premise that sets the plot in motion;
- that Miata is more serious about the election, an inference about Miata's motivation based on the promises each candidate has made;
- two units that describe the condition of the school, thus motivating Miata's goal to clean it up;
- a unit that describes how Dona Carmen will help her—by giving her flowers;
- a unit representing the final scene of Miata daydreaming about how the school will look.

Only two of the unique units have little relevance for the major chain of events (that Miata is related to the mayor and that the mayor ran against her husband in the election). Thus six of the eight unique units represent motivations, explanations of goals and intentions, and, finally, a recognition of how the story events might resolve. This suggests that the understanding that was developed based on the QtA lesson and discussion allowed students to build a representation of the story that included important connections and exhibited coherence.

The other narrative is *The Fun They Had*, an Isaac Asimov (2005) story about children in the future who are schooled at home by a computer, and is summarized below:

Margie and Tommy find a book about the “old days” and learn that children used to go to school together and learn from a human, and Margie imagines the fun the children must have had in those days.

As with other texts, the recalls for both the Strategies and QtA students have similar outlines. The prototypical recalls of both groups show that students typically recall that the story takes place in the future, when the children find a book about the past; they learn that school was taught by a human in contrast to Margie and Tommy's learning from a “robot,” and that Margie wishes she could go to one of those schools.

The QtA groups' recall includes six unique units. Most important among these are ones that relay that Margie and her peers now learn at home, by themselves, while the children all learned together in the old days. It is this contrast that causes Margie to think about “the fun they had.” The strategies groups' recall includes only one unique unit, that the children found the book confusing.

The expository text for which there were differences in recall was about the infrasonic

sounds that whales and elephants emit, *Messages by the Mile* (Facklam, 1992), which is described below.

The text begins by talking about how whales can communicate infrasonically and the scientists who have studied the whale sounds. The text then explains how that understanding led a scientist to discover that elephants can also communicate infrasonically.

Both groups' prototypical recalls include much of the basic information, such as that the text is about whales' and elephants' communication; that some of the sounds are infrasonic and can't be heard but can be felt as vibration, and that scientists study the sounds. The QtA group includes five unique units. Two play a key role in linking ideas within the text. These are that the scientists recorded the sounds (which the text explains cannot be heard but can be recorded) and the reason for doing so—that they could then figure out how whales communicate.

Examining Classroom Lesson Discussion

The patterns in the lesson recalls suggest that the lessons helped students in the QtA group to develop more connected understandings of the texts they read. We can explore how the lessons may have led to students' ability to produce such recalls through the transcripts of the lessons themselves. All of the lessons conducted during the study were recorded and transcribed. Below we present excerpts from two classrooms, one using the Strategies instruction and the other using QtA. The classes are reading and discussing *Off and Running* and are at the point where Miata's mother has told her about her relative, Dona Carmen, and Miata has decided to visit the woman. The important relationship that both teachers are scaffolding students to build is between Miata's wanting help with her election and Dona Carmen's experience winning elections as mayor of a Mexican town.

QtA Lesson Excerpt

In the excerpt, the teacher begins with an open-ended prompt to initiate discussion and invite students to make their thinking public. The teacher's goal is to have students recognize why Miata is seeking advice from Dona Carmen (who is her grandmother's sister-in-law and an ex-mayor) and to connect that idea to a previous section of text in which Miata expresses frustration about the upcoming election and is looking for "someone famous" to endorse her campaign. Several students weigh in and contribute to building the major idea of the text segment.

Teacher: So what just happened? What just happened? Elane.

Elane: She just found out that her sister was a mayor like she wants to ask her about how to win the elections.

Scott: It's her sister-in-law.

Teacher: Okay.

Lea: Her grandmother's sister-in-law was the mayor, I think.

Teacher: Greg.

Greg: She, she wanted to like, she wanted, she um, she said if she wants somebody famous and then she thought, like she thought um, her um, her grandmother's sister-in-law with, her grandmother's sister-in-law would, would help her go to her school and then do them do the election because she has experience because

she, I think she's done, she done experience with elections because she had to get ready too.

Teacher: So how does that connect to what we talked about earlier? I think you said it but how does that connect to what we talked about earlier? Antonio.

Antonio: Um, like if she, if she got advice from a real mayor, she would like, be able, she didn't know what, she didn't know what the main things she would need to do and how to handle them and like then that would get her more votes, more votes and basically that connects to what, to what we went over um before.

Strategies Lesson Excerpt

The following is the excerpt from the Strategies lesson that occurred at the same stopping point as the above QtA excerpt. In this excerpt, the teacher initiates the discussion by asking the students to summarize. When students are unable to respond to this initial prompt, the teacher directs their attention toward the process involved in developing a summary. The guides of *who*, *what*, *when*, *where*, *why*, and *how* are offered, and students contribute various components with some teacher guidance.

Teacher: Let's stop and summarize what we just read. Who would like to summarize what was just read? {No student response} What do you do when you summarize? What do you need to do? What is summarizing, Tina?

Tina: The *who*, *what*, *when*, *where*, *why* and *how* of the story.

Teacher: Okay, we take what we just read and tell the most important parts in our own words answering *who*, *what*, *where*, *when*, *why* or *how*. Does that help, Rochelle? Okay, let's try to work on that then. Coming up with a good summary of what was read. Let's try answering the *who*. Who are we talking about? Peter.

Paul: We're talking about Miata.

Teacher: Good. The *who* is Miata. What about the *what*? Sonia.

Sonia: She's picking up the bread.

Teacher: She's picking up bread. Would that be the most important part of what we just read?

Emily: No.

Tonia: She was going to the woman's house, the former mayor that her mother knew.

Teacher: Okay. That I like. Tonia just said the *what* is that she's going to the former mayor that her mother knows; the former mayor's house that her mother knows, Dona Carmen, is what Sonia said. To her house. Yeah, that's more important, I think, than the fact that she might be stopping to pick up bread also. So we want to make sure that we, we leave out the details but we keep in the important parts so the *who* is Miata, the *what* is she's going to meet a former mayor that her mother knows. When? When is this happening? Michelle.

Michelle: After they eat breakfast.

Teacher: Good. After they eat breakfast; what day did we say this was?

Derek: Saturday.

Teacher: Good. Saturday morning. Okay. Um, where? Let's answer the *where*. Paul.

Paul: Um, Dona Carmen's house.

Teacher: Yes, very good. It's at the woman's house, Dona Carmen's house. And why? I think that's important. Why is this happening? Who is the why of what you just read? Madelaine.

Madelaine: For her campaign.

Teacher: Okay, for her cam . . . the *why* is for her campaign. Who's the her?

Madelaine: Miata.

Teacher: Miata . . . Um, but why is she going to the former mayor, Dona Carmen's house?

Madelaine: So, so she can help her with her campaign.

Teacher: Okay. Okay. Good. The *why* is Dona Carmen may be able to help her with her campaign. That was good. Um, with everything we just said, was that a good summary?

Kevin: Yes.

Teacher: Ok. Li. What made it a good summary? Why can we say yes, that was a good summary? Shall I put it all together? Would that help? Okay. Um, Miata is going to meet a former mayor that her mother knows on Saturday morning at the woman's house because this woman may be able to help Miata with her election to become class president. Is that a good summary? Can you tell me why now? What makes it good?

Li: Because we answered the important parts.

Teacher: The important, uh, parts of what we just read were answered and what helped us again? Let's recap. What helped us make that good summary? What did we use? Michelle.

Michelle: Um, *who, what, when, where, why, how.*

The QtA and Strategies excerpts presented reflect two systematic patterns we found across the transcripts. The first is that in QtA lessons a greater proportion of the talk was directly about text content. We examined the transcripts for four lessons for all classrooms. The lessons examined were those for the four texts—two narratives and two expository—for which students gave recalls. In QtA lessons, 94% of student talk and 50% of teacher talk was about the text, while in Strategies lessons, 75% of student talk and 27% of teacher talk was based on the text.

In the Strategies lessons, more talk was taken up by referring to the strategies and how to implement them. But that did not end up translating into better comprehension, or in transfer of skill to comprehension of text on their own (the transfer task).

The second pattern is that students produced many more words per turn in the QtA lessons than in the Strategies lessons. At this point, we have only analyzed the two narratives for this pattern, but it is quite strong, with QtA student turns averaging 24 words relative to 12 in Strategies lessons. More focus on the text in discussion and taking opportunity to talk at some length may underlie the comprehension advantage evidenced in the recalls. In particular, the greater number of connections included, as demonstrated in the prototypical recalls, may have come about as students gave more language to their understandings during discussion, thereby including more connections. It could be argued that in producing more language about text and more words per turn, the QtA students were simply responding to the teacher's expectations about participating in a discussion. Our view, however, is that responding to such expectations encourages students to build an implicit model of what comprehension entails. Participating in discussions framed around open questions and encouragement to elaborate and connect ideas shapes the way students deal with and comprehend text.

What is the Relevance of our Findings to Metacognition?

In this section we offer a perspective on metacognition and consider what our findings suggest about how metacognition can be used to support students' reading. It seems that strategies instruction and QtA take different approaches to metacognition. The way the two evince metacognition parallels discussion in the literature about what makes actions metacognitive versus cognitive. This discussion is reflected in Hacker's (1998) argument for the convention that metacognition be reserved for conscious and deliberate thoughts that are not only potentially controllable, but also potentially reportable. On the other hand, Gavelek and Raphael (1985) suggest that much of metacognition may be tacit. They contend that the utility of metacognition for instruction resides not in students' knowledge of their cognition, but in how it affects their performance. Gavelek and Raphael see metacognition as going "to the heart of the cognitivists' assumption of the learner as an active organism" (p. 129). Their implication here seems to be that knowledge that enables learners to influence their mental activity is metacognitive, even if that knowledge does not involve a deliberate, conscious decision.

A suggestion by Sinatra et al. (2002) also highlights the importance of this active stance in contrast to the explicit knowledge of strategy application. Sinatra et al. suggest providing students with experience responding to the kinds of questions that can lead them to take "an active, strategic stance toward text" (p. 70) as an alternative to strategies instruction. They explain their choice by saying that allocation of deliberate attention while reading may not involve deliberate awareness—competent readers don't consciously tell themselves "pay attention to what's important." Further, Sinatra et al. consider that continued attention to deliberate use of strategies may undermine comprehension. This is because comprehension takes significant mental resources, which are limited. So if some resources are devoted to calling up strategies, adequate resources may not be directed toward the actions needed for comprehension, because students are being asked to do something in addition to making sense of texts.

Our findings support the notion that the key to getting students to take control of their own processes while reading may not necessitate knowledge of and conscious attention to specific processing actions. Rather it may involve deliberate, but not conscious, attention to text content in ways that promote attending to important ideas and establishing connections between them. QtA does not ask for specific knowledge of what the processes entail, but of knowledge that there *is* a process—that readers need to apply effort and deliberately try to figure things out as they read. Students get this message about their process over and over as they are provided regular scaffolded practice during reading. That is, the teacher initiates deliberate consideration of text with open questions, and follows up with prompts to elaborate, reconsider, and connect. This guidance and the ensuing discussion seem to help students learn to regulate their actions as they read. Practice with the QtA approach, according to our findings, can lead to enhanced comprehension, including of text that students read on their own. Most particularly it is the connections in text that seem to be enhanced.

It may well be that strategies instruction helps make students active, too, as our results show that students experiencing the Strategies approach showed overall competent comprehension. But strategies instruction brings about active processing by way of bringing processes to conscious awareness. A question, then, is why not go directly to an active stance without the explicit invoking of specific routines? Our findings suggest that adding this into the mix does not enhance comprehension, and it consumes time and attention. In addition, as judging from the prototypical recalls, working toward metacognition through strategies may leave students especially vulnerable to missing connections within a text.

Albeit we are hypothesizing based on limited information. Ours was a small-scale study and the differences that we found are not large. However, results did show a consistent trend from the lesson discourse through recalls of both lesson text and transfer text. And the trend of greater recall, especially in regard to connections in text, is explicable in terms of specific aspects of cognitive processing theory. One aspect is the importance of connections in a reader's building meaning from text; QtA questions and discussion explicitly focus on how information in a text connects. Another aspect relates to resource allocation during reading. That is, QtA's design that encourages readers to focus directly on text content doesn't put an extra burden on readers' limited mental resources.

References

- Anderson, V. & Roit, M. (1993). Planning and implementing collaborative strategy instruction for delayed readers in grades 6–10. *Elementary School Journal*, 94, 121–137.
- Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly*, 22, 331–346.
- Asimov, I. (2005). The Fun they Had. In *Trophies Distant Voyages* (pp. 584–592). Chicago: Harcourt.
- Baker, L. (1979). Comprehension monitoring: Identifying and coping with text confusions. *Journal of Reading Behavior*, 11, 366–374.
- Baker, L. & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of Reading Research* (pp. 353–394). New York: Longman.
- Baron, J. (1985). *Rationality and intelligence*. Cambridge, England: Cambridge University Press.
- Beck, I. L. & McKeown, M. G. (1998). Transforming knowledge into tangible resources to support pedagogical change: Final report to the Spencer Foundation.
- Beck, I. L. & McKeown, M. G. (2006). *Improving comprehension with Questioning the Author*. New York, Scholastic.
- Beck, I. L., McKeown, M. G., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the author: A yearlong classroom implementation to engage students with text. *The Elementary School Journal*, 96:4, 385–414.
- Beck, I. L., Omanson, R. C., & McKeown, M. G. (1982). An instructional redesign of reading lessons: Effects on comprehension. *Reading Research Quarterly*, 17, 462–481.
- Block, C. C. (1993). Strategy instruction in a literature-based program. *The Elementary School Journal*, 94, 103–120.
- Brown, A., Armbruster, B. B., & Baker, L. (1986). The role of metacognition in reading and studying. In J. Orasanu (Ed.), *Reading comprehension: From research to practice* (pp. 49–75). Hillsdale, NJ: Erlbaum.
- Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology*, Vol. III, *Cognitive development* (pp. 77–166). New York: Wiley.
- Brown, A. L., Campione, J. C., & Barclay, C. R. (1979). Training self-checking routines for estimating test readiness: Generalization from list learning to prose recall. *Child Development*, 50, 501–512.
- Brown, A. L. & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behaviour*, 22(1), 1–14.
- Brown, A. L. & Smiley, S. S. (1977). Rating the importance of structural units of prose passages: A problem of metacognitive development. *Child Development*, 48, 1–8.
- Carver, R. P. (1987). Should reading comprehension skills be taught? In J. E. Readance & R. S. Baldwin (Eds.), *Research in literacy: Merging perspectives* (Thirty-sixth yearbook of the National Reading Conference, pp. 115–126). Rochester, NY: National Reading Conference.
- Duffy, G. G. & Roehler, L. R. (1989). Why strategy instruction is so difficult and what we need to do

- about it. In C. B. McCormick, G. E. Miller, & M. Pressley (Eds.), *Cognitive Strategy Research* (pp. 133–157). New York: Springer-Verlag.
- Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., Vavrus, L. G., Wesselman, R., Putnam, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, 22(3), 347–368.
- Facklam, M. (1992). Messages by the Mile. In *Bees Dance and Whales Sing* (pp. 35–40). San Francisco: Sierra Club Books for Children.
- Garner, R. (1981). Monitoring of passage inconsistency among poor comprehenders: A preliminary test of the “piecemeal processing” explanation. *Journal of Educational Research*, 67, 628–635.
- Garner, R. & Kraus, C. (1981–1982). Good and poor comprehender differences in knowing and regulating reading behaviors. *Educational Research Quarterly*, 6, 5–12.
- Gavelek, J. R. & Raphael, T. E. (1985). Metacognition, instruction, and the role of questioning activities. In D. L. Forrest-Pressley, G. E. MacKinnon, & T. Gary Waller (Eds.), *Metacognition, Cognition, and Human Performance* (pp. 103–136). New York: Academic Press, Inc.
- Gersten, R., Fuchs, L., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities. *Review of Educational Research*, 71(2), 279–320.
- Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1–23). Hillsdale, NJ: Erlbaum.
- Jenkins, J. R., Heliotis, J., Stein, M. L., & Haynes, M. (1987). Improving reading comprehension by using paragraph restatements. *Exceptional Children*, 54, 54–59.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction integration model. *Psychological Review*, 95, 163–182.
- Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394.
- Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. *The Elementary School Journal*, 99(1), 3–20.
- Markman, E. M. (1979). Realizing you don’t understand: Elementary school children’s awareness of inconsistencies. *Child Development*, 50, 643–655.
- McKeown, M. G. & Beck, I. L. (1994). Making sense of accounts of history: Why young students don’t and how they might. In G. Leinhardt, I. L. Beck, & K. Stainton (Eds.), *Teaching and learning in history* (pp. 1–26). Hillsdale, NJ: Erlbaum.
- McKeown, M., Beck, I., & Blake, R. (in press). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*.
- McKeown, M. G., Beck, I. L., & Sandora, C. A. (1996). Questioning the author: An approach to developing meaningful classroom discourse. In M. G. Graves, B. M. Taylor, & P. van den Broek (Eds.), *The first R: Every child’s right to read* (pp. 97–119). Teachers College Press.
- McKeown, M. G., Beck, I. L., & Worthy, M. J. (1993). Grappling with text ideas: Questioning the author. *The Reading Teacher*, 46(7), 560–566.
- Miller, G. E. (1985). The effects of general and specific self-instruction training on children’s comprehension monitoring performances during reading. *Reading Research Quarterly*, 20(5), 616–628.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific literature on reading and its implications for reading instruction*. Washington, DC: National Institutes of Health (NIH Pub. No. 00–4754).
- Olshavsky, J. (1976–1977). Reading as problem solving: An investigation of strategies. *Reading Research Quarterly*, 12, 654–674.
- Omanson, R. C. (1982). An analysis of narratives: Identifying central, supportive, and distracting content. *Discourse Processes*, 5(3–4), 195–224.
- Omanson, R. C., Beck, I. L., Voss, J. F., & McKeown, M. G. (1984). The effects of reading lessons on comprehension: A processing description. *Cognition and Instruction*, 1, 45–67.
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 2, 117–175.

- Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology*, 76, 1239–1252.
- Pearson, P. D. & Dole, J. A. (1987). Explicit comprehension instruction: A review of research and a new conceptualization of instruction. *The Elementary School Journal*, 88(2), 151–165.
- Pearson, P. D. & Fielding, L. (1991). Comprehension instruction. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research*, Vol. 2 (pp. 815–860). New York: Longman.
- Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, 92(5), 513–555.
- Rinehart, S. D., Stahl, S. A., & Erickson, L. G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly*, 20(4), 422–438.
- Rosenshine, B. & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479–530.
- Royer, J. M., Hastings, C. N., & Hook, C. (1979). A sentence verification technique for measuring reading comprehension. *Journal of Reading Behavior*, 11, 355–363.
- Sandora, C., Beck, I., & McKeown, M. (1999). A comparison of two discussion strategies on students' comprehension and interpretation of complex literature. *Reading Psychology*, 20(3), 177–212.
- Schmitt, M. C. (1988). The effects of an elaborated directed reading activity on the metacomprehension skills of third-graders. *National Reading Conference Yearbook*, 37, 167–181.
- Sinatra, G. M., Brown, K. J., & Reynolds, R. (2002). Implications of cognitive resource allocation for comprehension strategies instruction. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 62–76). New York: Guilford.
- Smiley, S. S., Oakley, D. D., Worthen, D., Campione, J. C., & Brown, A. L. (1977). Recall of thematically relevant material by adolescent good and poor readers as a function of written versus oral presentation. *Journal of Educational Psychology*, 69, 381–387.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.) (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Soto, G. (2005). Off and Running. In *Trophies Distant Voyages* (pp. 492–506). Chicago: Harcourt.
- Sternberg, R. J. (1979). The nature of mental abilities. *American Psychologist*, 34, 214–230.
- Sternberg, R. J. (1982). A componential approach to intellectual development. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*, Vol. 1. Hillsdale, NJ: Erlbaum & Associates.
- Symons, S., Snyder, B. L., Cariglia-Bull, T., & Pressley, M. (1989). Why be optimistic about cognitive strategy instruction? In C. McCormick, G. Miller, & M. Pressley (Eds.), *Cognitive strategy research: From basic research to educational applications*. New York: Springer-Verlag.
- van den Broek, P. (1990). The causal inference marker: Towards a process model of inference generation in text comprehension. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 423–446). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wagoner, S. A. (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly*, 18(3), 328–346.
- Winograd, P. & Johnston, P. (1987). Some considerations for advancing the teaching of reading comprehension. *Educational Psychologist*, 22(3 & 4), 213–230.

3 The Role of Metacognition in Teaching Reading Comprehension to Primary Students

Joanna P. Williams and J. Grant Atkins
Teachers College, Columbia University

Learning to read with understanding is the most important achievement in a young student's life. Unfortunately, many students are unable to comprehend texts, even though they can decode them fluently. Consequently, teachers, researchers, and other educators have begun focusing on early comprehension instruction. This chapter addresses the role of metacognition in teaching reading comprehension to primary students. It includes a review of studies on the development of metacognition and a review of instruction designed to develop metacognition both in young children and in teachers. We begin with a short discussion of some of the basic concepts relevant to the study of comprehension and comprehension instruction. Variation in the way these concepts are defined sometimes leads to confusion.

The RAND Study Reading Group defined reading comprehension as the extraction and construction of information through the involvement and interaction with a text (RAND Study Reading Group, 2002). Successful comprehension involves sifting through a text to identify its main point and often going beyond that to critically evaluate or apply what has been understood. It consists of many components, such as identifying word meanings, processing sentences, linking ideas across sentences, and inferencing. These components can be analyzed into myriad subcomponents and subprocesses; theories of comprehension (Graesser, 2007) do just that.

Many of the concepts central to these theories are difficult to define. *Skills* are those competencies that a reader brings to a text, such as decoding or inferencing abilities. Ideally, a reader has the skills that allow all the processes necessary for comprehension to work in concert, quickly and effortlessly, without reaching conscious awareness. Of course, reading does not always follow this idealized path. Sometimes even proficient readers stumble over unfamiliar vocabulary or abstruse technical explanations. When this happens, comprehension breaks down, and readers must employ conscious *strategies* to repair their comprehension. They might re-read a portion of the text, or they might ask themselves questions to highlight certain information in the text. In some reading situations, comprehension is quickly repaired and strategies go by unnoticed; other situations require deliberate effort.

Over the past few decades, a consensus has emerged that these strategies are at the heart of what should be taught to students to improve their comprehension (National Reading Panel, 2000; Pressley, 2005). Comparison of good and poor students suggested that poor readers often do not perform repair strategies when necessary. Indeed, they may even lack the ability to notice the inadequacy of their comprehension. These readers take a passive approach (Haines & Torgesen, 1979), and when their comprehension fails, they are not able to cope.

Current reading instruction focuses on helping readers learn and use strategies that will improve their comprehension. The goal is to enable students to internalize and automatize

these strategies, turning them into skills (Samuels, 2002). While the distinction between skills and strategies is easily made in the abstract, in practice it is difficult to differentiate one from the other. Educators use the terms in a variety of ways, and in some cases one would be hard pressed to identify a particular aspect of a student's performance as either skillful or strategic.

The difference between *cognition* and *metacognition* is another important theoretical distinction. Metacognition can be broadly defined as cognition about one's own cognitive processes (Flavell, 1979; Baker, 2002). Most definitions of metacognition have focused on two separate but related aspects: (1) knowledge/awareness of cognitive processes, and (2) control of cognitive processes. The first aspect can be further subdivided into the knowledge that people experience cognitions (theory of mind) and the awareness of one's own cognitive processes as they relate to tasks and to other people. The second aspect of metacognition can also be broken down into two components, the monitoring of cognitive processes (knowing when they are and are not being used effectively) and the ability to regulate cognition to improve effectiveness (using strategies to repair comprehension failures, for example).

The distinction between cognition and metacognition also makes sense in the abstract, but it is extremely slippery. It is difficult if not impossible to categorize a particular reading activity as wholly metacognitive or not-at-all metacognitive. Moreover, some curriculum designers may describe their instructional programs as fostering metacognitive strategies while other curriculum designers may use largely the same set of instructional techniques but characterize them as promoting cognitive strategies.

A perusal of the recent literature suggests that as important as these distinctions may be theoretically, insisting on clear differentiations is probably less productive when it comes to working on instructional applications. As Duffy (2005) points out, "In reading instruction, metacognition is associated with reading strategies." This seems clear enough for our purposes. In this chapter, we describe studies using the same language as the researchers who conducted the studies. Whatever the terminology, our focus is always on awareness and control of cognitive processes.

The study of metacognition was introduced by Flavell (1976), whose initial focus was on the development of children's memory. Flavell traced the course of acquisition of the ability to reflect on and control one's own memory processes. As they get older, children develop the ability to use strategies such as active rehearsal, organization into categories, and, later, elaboration (Kreutzer, Leonard, & Flavell, 1975). At a certain point, children become aware of their own memory processes and can begin to control them, by deliberately rehearsing the information they wish to remember or by organizing the information into categories. Some of this cognitive activity is done at an automatic level, as when a child, asked to remember a word list, recites the words in categories without any intention or realization of doing so. But when a memory task is more difficult, a child may apply a strategy with effort. At this point, he becomes conscious of what he is doing, and he is using a metacognitive strategy.

Flavell's early work inspired two important strands of research, which up to now have remained rather separate. One, called theory of mind research, deals with very young children—toddlers and pre-schoolers. The other is currently centered on applications of metacognitive theory to instructional issues. Until recently, most of the second type of research has been focused on children at the fourth-grade level or older. Currently, there is some interest in looking at primary-level children.

Theory of Mind: The Precursor of Metacognition

The developmental progression of metacognition from awareness to regulation of cognitive processes is clear; after all, it is impossible to regulate something unless one first possesses it. It is less clear *how* these competencies develop; that is, what are the prerequisites for the knowledge and regulation of cognition and which environments and instruction can assist in their development?

Investigation of the earliest stages of metacognition is known as the study of the theory of mind (Flavell, 2000; Kuhn, 2000). In short, the development of the theory of mind is the precursor to the development of the first part of metacognition: knowledge of cognitive processes. To have a theory of mind means to be aware that one has knowledge and beliefs that are shaped by one's experiences and that other people's experiences shape their knowledge and beliefs. This awareness of cognition, and the separation of cognition from perception, typically occurs around the age of 3 (Kuhn, 2000).

The development of theory of mind can be assessed by a number of paradigmatic tasks that tap this understanding. A classic research paradigm is the false belief task. A child can be said to have a sense of belief (i.e., of a mental state) and therefore a theory of mind if the following scenario holds: the child sees someone put an object in a box and then leave. The child then sees someone else remove the object from the box and put it into a second box. If the child expects that the first person he observed, upon returning, will look for the object in the first box, he can be said to have a theory of mind. That is, he understands that people can hold beliefs that are contradictory to reality, that there is a distinction between the mind and the world (Wellman, Cross, & Watson, 2001).

In addition to the false belief test, other tasks such as perspective-taking and distinguishing between appearance and reality have been used to demonstrate that very young children grow out of their early egocentrism and acquire a "theory of mind," such that they can think about mental states, and, as they get older, develop the knowledge that they can examine and control their own cognitive abilities. Once children have developed a theory of mind they have the infrastructure necessary to develop cognitive and metacognitive strategies. In other words, theory of mind is a prerequisite for metacognition, and those experiences and traits that lead to the development of a theory of mind also lead to the development of metacognition. In a sense, metacognition is the practical application of theory of mind to cognitive tasks; theory of mind provides the conceptual underpinnings needed to develop and use metacognitive knowledge (Lockl & Schneider, 2006). Children who see themselves and others as people who are influenced by their mental states and act upon them can then reflect on (and eventually regulate) these mental processes.

Language is closely related to the development of theory of mind. Most of the relevant studies of this relationship focus on the role of basic language skills in accelerating the development of the theory of mind (see Milligan, Astington, & Dack, 2007, for a meta-analysis of such studies). Other studies cited in this meta-analysis showed that a child's exposure to and use of metacognitive terms (e.g., think, know, believe) are also associated with a more rapidly developing theory of mind (Milligan, Astington, & Dack, 2007). Thus research suggests that talking and hearing others talk about cognitive processes helps establish knowledge of these processes.

Recent longitudinal studies have shown that the relationship between language and theory of mind is bidirectional. Slade and Ruffman (2005) tested the abilities of 44 children with a mean age of 3.8 years (at first testing) in theory of mind, language, and working memory at two time points separated by 6 months. Theory of mind was assessed through a series of false belief tasks. Language was assessed using two semantic tests (vocabulary and linguistic concepts) and two syntactic tasks (word order and

embedded clauses). Working memory was also measured, through a modified backwards digit span task. After establishing the stability of the constructs and equating the sensitivity of the language and theory of mind tasks, Slade and Ruffman found that not only does language ability facilitate the development of theory of mind, but theory of mind can also facilitate the later acquisition of language. Working memory did not account for this relationship. Further, they found that both syntax and semantics play a significant role in theory of mind acquisition.

A longitudinal study by Lockl and Schneider (2006) examined theory of mind, metamemory, and language, focusing on the comprehension of metacognitive vocabulary. Lockl and Schneider assessed children four times at 6-month intervals, with the first assessment at approximately 4.5 years of age. The first testing consisted of three theory of mind tests (two false belief tasks and an appearance-reality task in which children answered questions about an object that looked like a different object). At the three subsequent assessments, metamemory, metacognitive vocabulary, and general vocabulary were tested. Metamemory was assessed through an interview in which the participants were presented with a description of the memory strategies of two different children and had to pick which child would have better memory. Metacognitive vocabulary was assessed by reading to the participants a story accompanied by pictures and then asking them about the character's mental state.

Performance on the theory of mind tasks at age 4.5 predicted metamemory, metacognitive vocabulary, and general vocabulary up to a year and a half later. The correlations between metacognitive vocabulary and metamemory increased between the ages of 5 and 6. These two constructs showed a bidirectional relationship; that is, early metacognitive vocabulary predicted later metamemory, and early metamemory predicted later metacognitive vocabulary. Lockl and Schneider (2006) proposed that the acquisition of specific metacognitive vocabulary enables a child to think about his own memory and figure out what would enhance memory performance. Likewise, increased metacognitive knowledge of memory can aid the comprehension of metacognitive vocabulary.

Although there is some research on more “advanced” theory of mind in older autistic children and adults (Baron-Cohen, 2001), theory of mind research typically focuses on pre-school students and has not been of great interest to education researchers. However, as the latter become more interested in providing comprehension instruction to primary-level students, they may seek to learn more from this area of research. We come back to this point later in this chapter.

The Development of Metacognition in Young Children: Empirical Studies of Reading

The concept of metacognition was initially applied to the field of reading by Brown (1980), who described the reading process as involving strategic knowledge and action: smooth sailing—comprehension at an automatic level—until comprehension breaks down, and then conscious attempts to comprehend via re-reading, looking at pictures, figuring out meanings of unknown words, parsing sentences, etc. Brown proposed that a lack of metacognitive processing is the reason why many children are not successful readers. Since her analysis appeared, most of the research in reading comprehension has revolved around comprehension strategies: what they are, how they operate in ordinary reading, and what to do when they do not function well. This focus is also prevalent in research on the writing process (Griffith & Ruan, 2005). We have limited our review to studies that deal with preschool and early primary-grade students and do not review the much larger body of research involving older children.

Within reading research, metacognition has traditionally been viewed as a late-developing competency. For many years, educators and researchers believed that young children did not have metacognitive knowledge or skills and that metacognitive instruction was not only a waste of time, but quite possibly detrimental to a child's learning. One important objection to metacognitive instruction in reading is based on the concept of executive functioning, which involves the coordination of various cognitive processes to accomplish a task. Given the fact that executive functioning has a limited capacity, many educators believed that lower-level skills, such as phonological awareness and decoding, need to be mastered and automatized before executive functioning can devote resources to higher-level skills, such as metacognition.

Empirical studies, however, have demonstrated that children as young as 4 years old display metacognitive knowledge and strategies while reading. The true extent of metacognition is difficult to determine as children may possess knowledge and use strategies that they are unable to express (Juliebö, Malicky, & Norman, 1998). The early work on the development of metacognition in young children relied primarily on interview data (Paris & Jacobs, 1984). Thus, our understanding of metacognition in young children was, and—to the extent that many studies still use interviews—is, limited by the assessment methods most frequently used in research (Paris & Flukes, 2005). The observed differences in the metacognitive knowledge and strategies of younger and older readers, therefore, may not necessarily be true differences in metacognition, but rather variations in the ability to describe cognitive and metacognitive processes. While most studies rely on methods that involve students' verbal expression, some studies have attempted to assess the metacognition of young readers using methods that are not completely dependent on interviews.

Brenna (1995) conducted a case study of five fluent readers between the ages of 4 and 6 and found that they employed a variety of metacognitive strategies while reading. She observed the children reading unfamiliar books, interviewed them during and after reading, recorded the types of errors they made, and administered a role play in which the children discussed reading with puppets. It is important to note that the five participants in this study were not representative of children their age, as they were well ahead of their peers in reading ability and had home environments where reading was clearly valued. Nevertheless, these children do provide an example of efficient metacognition in young readers. While reading, the children displayed metacognitive strategies that were based on self-knowledge, task-knowledge, or text-knowledge. The two most fluent readers used strategies that combined semantic, syntactic, and phonological cues when their comprehension broke down. By contrast, the least fluent reader relied primarily on sounding out words (phonological cues) when she faced difficulties. Parental reports indicated that the more fluent readers had used phonological strategies to repair comprehension before they learned other strategies. Thus, a developmental model progressing from primary reliance on phonological strategies to use of a wider repertoire of strategies is apparent. Not surprisingly, the children used those strategies that their caregivers suggested and modeled most often. In addition to using a range of strategies, the students responded to the various methods used to detect metacognitive behavior in different ways. Some children exhibited and discussed strategies during the read-aloud and interview portions of the study, whereas others demonstrated knowledge of metacognitive strategies only during the role play (Brenna, 1995).

The study by Juliebö et al. (1998), mentioned earlier, also demonstrated the importance of using a variety of assessment methods when studying metacognitive knowledge in young children. These researchers examined metacognition in five first-graders whom the teachers identified as having reading difficulties, i.e., performance on literacy tests

indicated that these children were at an emergent stage of literacy. The children were given an intervention program consisting of 14 to 16 weeks of daily 30-minute videotaped sessions. Instances of metacognitive behavior on the videotapes were identified. Selected videotapes were then shown to the children who answered questions designed to help the researchers retrospectively to identify the children's metacognitive processes. This procedure gave the children an opportunity to reflect on cognitive processes in reading without the confounding factor of memory. Interestingly, the first-graders were more likely to report an awareness of being right or wrong and to self-correct during the actual intervention, whereas they were more likely to display knowledge of comprehension and repair strategies during the retrospective videotaped sessions (Juliebö et al., 1998).

Phonics and use of pictures to identify words were the strategies most frequently identified in the retrospective sessions. This is consistent with Brenna's finding that less fluent readers primarily use sounding-out strategies in solving comprehension problems. Most of the children did not report using more than one strategy at a time; this is consistent with other studies showing that beginning readers have yet to integrate multiple strategies (Brenna, 1995). While the retrospective sessions demonstrated that the children had knowledge of metacognitive strategies, it was less apparent that they understood when and why to use the strategies (i.e., that they had acquired conditional knowledge). This lack of metacognitive regulation is reflective of Markman's classic studies (1977, 1979), which indicated that young children do not monitor their comprehension.

In a more extensive study of metacognition in beginning readers, Kinnunen, Vauras, and Niemi (1998) examined the comprehension monitoring processes of 132 Finnish first-graders (mean age, 7 years 10 months). This study provided evidence of comprehension monitoring even among students with poor decoding and listening comprehension skills. The authors used an online method of tracking reading speed and lookbacks (re-reading) at the sentence and passage level as indicators of comprehension monitoring. Across all students, comprehension monitoring was more apparent at the sentence than at the passage level, and also when the measure was reading speed as opposed to the number of lookbacks. Even poor decoders slowed down their reading when encountering semantic, syntactic, or factual knowledge violations at the sentence level. On the other hand, these poor decoders utilized repair strategies such as looking back and re-reading less often than average and good decoders. A similar pattern was observed when the children were divided according to listening comprehension skill. Good comprehenders were more consistent and effective monitors (Kinnunen, Vauras, & Niemi, 1998). These findings support the notion of a distinction in comprehension monitoring between knowledge of difficulties only in poor readers (demonstrated by slower reading) and knowledge plus strategic regulation to repair comprehension in stronger readers (demonstrated by lookbacks and re-readings).

In addition to these studies showing evidence of metacognition in young readers, there is also evidence of metacognition in young writers. Ruan (2004) investigated the metacognitive knowledge displayed by 16 bilingual Chinese/English first-graders using a dictation task developed by Cox (1994). This task required the children to dictate a story as a text for others to read. The task was designed to elicit metacognitive utterances related to the planning, regulating, and editing processes that are necessary for the dictation to conform to written conventions. Sessions were conducted at the beginning and the end of first grade, and taped sessions were analyzed for instances of declarative metacognitive knowledge (e.g., writing goals, text structure, and metalinguistic comments) and procedural metacognitive knowledge (e.g., planning, thinking, and regulatory comments). The children made significantly more procedural metacognitive comments at the end of first grade than at the beginning, but there was no change in the number of declarative

metacognitive comments from the beginning to the end of the year. A qualitative analysis revealed that procedural metacognitive comments included inner thinking (e.g., “. . . um . . .,” “I think. . .”), self-regulatory speech (e.g., “I am mixed up . . .,” “I mean . . .”), and other-regulatory speech (e.g., “Cross that out,” “Erase it”). It also revealed that the poor writers tended to make fewer metacognitive utterances, whereas “more advanced writers tended to comment more often on the task before, during, and after they dictated the story” (Ruan, 2004, p. 110). This study is noteworthy for its use of a novel tool for investigating metacognitive development in young writers. More studies are needed that utilize this dictation technique and other novel techniques that do not rely on children’s ability to explain their metacognitive knowledge.

These empirical studies lead to several conclusions. First, even young children possess and use metacognitive strategies while reading. Second, the choice of metacognitive strategies seems to depend both on the reader’s developmental level and the assessment method used. It appears that younger and less skilled readers tend to use phonological, or sounding-out, strategies when faced with comprehension difficulties, whereas older and more skilled readers also use semantic and syntactic cues to repair comprehension. While it seems clear that many young readers have knowledge of cognitive strategies and a metacognitive awareness of when their comprehension fails, it is less obvious that they are able to regulate metacognitive strategies in order to repair comprehension. More sensitive assessment methods and larger studies are needed to validate these conclusions.

Metacognitive Reading Instruction for Young Children

Some Highlights of the Research to Date

The National Reading Panel (2000) reviewed 205 studies that evaluated the effectiveness of teaching text comprehension. A small proportion of these studies dealt with primary-age children. Sixteen categories of instruction were identified in the review. There was a solid scientific basis for concluding that seven of these improve comprehension in non-impaired readers: comprehension monitoring, cooperative learning, use of graphic and semantic organizers, question answering, question generation, story structure, and summarization. Most of the reviewed studies evaluated the effectiveness of instruction in a single cognitive strategy. A smaller number of more recent studies evaluated the effectiveness of instruction in small sets or “packages” of strategies, which replicate more closely what actually goes on in a classroom, where single strategies rarely appear in isolation. No one strategy is always effective, and it is only through learning a number of flexible strategies that students can become metacognitively aware of the effectiveness of specific strategies in specific situations (Baker 2002). The corpus of studies in the NRP meta-analysis did not lend itself to attempts to distinguish between cognitive and metacognitive elements in the strategy research that was examined.

A study by Glaubman, Glaubman, and Ofir (1997) looked specifically at the use of a metacognitive method of instruction and provided evidence of the effectiveness of teaching metacognitive strategies through an investigation of question generation by kindergarteners. They taught children to generate questions according to three methods, one based on metacognitive theory, one based on active processing theory, and one conventional method. The metacognitive method focused on raising the children’s awareness of the processes involved in questioning during stages of learning question words (*what, why, how, etc.*), matching questions to knowledge, and understanding the purposes of questions. At each stage, the students were taught to think about the questions they asked and how the answers to these questions increased their knowledge. Students taught by the

active processing method participated in activities designed to encourage the generation of questions; the focus was to improve questioning skills and increase questioning vocabulary by generating as many and as varied questions as possible. In the conventional method, students were encouraged to generate questions throughout the day and in different parts of the curriculum, but there was no explicit instruction, and no specific portion of the day was set aside for training in question generation. Kindergarteners participated in the intervention for 15 weeks, with 30 minutes of instruction each week. The pre-test and post-test consisted of three measures, quality of questions (measured by categorizing questions elicited in response to seeing a hamster in a cage [pre-test and post-test] and an African statue [post-test]), story comprehension (measured by comprehension questions, sequence picture arrangement, and verbal recall), and self-directed learning (measured by observations of a problem-solving activity). The metacognitive training group fared better than the active processing or conventional groups on all three measures on the immediate post-test. On a three-month delayed post-test, the metacognitive group scored higher than the other two groups on the question quality measure. However, there was no difference between metacognitive and active processing groups on story comprehension. Self-directed learning was not assessed on this delayed post-test.

The results of the Glauberman et al. (1997) study demonstrate the value of integrating metacognitive strategies into instruction focused on question generation in young children. They suggest that instruction in metacognitive awareness, in addition to typical instruction in reading and questioning strategies, can develop self-directed and regulated learners. Metacognitive training helps children internalize the strategies they use and promotes an awareness of when and why they are effective. This awareness bolsters the ability to transfer the strategies (as measured by their ability to ask questions about a novel object) to other situations in which they would be useful.

Instructional Programs in Reading Comprehension

Positive evidence of the effectiveness of metacognition instruction has contributed to the design of several broader instructional programs that are most often strategies-based. Sometimes the strategies are classified as metacognitive, while at other times they are considered cognitive strategies that are monitored and regulated metacognitively. Most of the instructional programs explicitly model the strategies and then provide scaffolding to ensure that the students understand and can use the strategies effectively. Some of them were designed for older students and then adapted for primary-level students.

Informed Strategies for Learning (ISL), developed by Paris and colleagues (Paris, Cross, & Lipson, 1984), was one of the first programs to be developed and demonstrated that metacognitive strategy instruction is feasible in the primary grades. ISL focuses on providing students with declarative (what strategies are), procedural (how to use them), and conditional (when are they most effective) knowledge about reading strategies. The procedural and conditional components of ISL made Paris a forerunner in understanding that students need knowledge of when and why to use reading strategies in order to implement them effectively. These strategies were taught to third-graders using explicit instruction, metaphors and visual images on bulletin boards, and information for teachers in how to incorporate the strategies into other areas of the curriculum. The instruction was provided in three stages: importance of strategies, specific strategies to use while reading, and comprehension monitoring. ISL instruction resulted in what has become a typical outcome for studies examining metacognitive strategy use: a significant increase in strategy knowledge and improvement on experimenter-developed comprehension measures, but no difference in performance on standardized tests. Thus, ISL students made significantly

greater gains on specifically constructed comprehension measures (cloze and error detection tasks) than non-ISL students and had more knowledge and awareness of cognitive and metacognitive comprehension strategies. However, this knowledge and awareness did not translate into improved performance on the Gates-McGinitie Reading Tests or the Tests of Reading Comprehension, where there was no significant difference between the ISL and control groups. It is not unusual in such studies to find differences on experimenter-developed tests but not on standardized tests. In most cases the former have been designed to focus specifically on the topics covered in the instruction, whereas standardized tests are likely to encompass a broader range of topics.

Reciprocal Teaching (RT; Palincsar & Brown, 1984) is one of the most researched and widely used programs for teaching comprehension strategies. The two main features of the program are its focus on teaching four comprehension strategies (summarization, question generation, clarification, and prediction) and that teaching is structured as a dialogue between the teacher and students. This dialogue includes modeling of the strategies, elaboration on students' use of the strategies, assistance in their use, and feedback. Students are encouraged to participate in the dialogue until they can assist other students without the teacher's help. A meta-analysis by Rosenshine and Meister (1994) examined 16 studies that evaluated RT. Studies that used a standardized test (11 studies total, two of which yielded significant differences) had a moderate effect size, whereas studies that used experimenter-developed tests (either short-answer or summarization tests; 10 studies total, eight with significant differences) had a large effect size. Five studies assessed the students using both a standardized test and experimenter-developed tests, and four of them yielded significant results on the experimenter-developed tests but not on the standardized tests.

In addition, Rosenshine and Meister (1994) divided the studies according to ability level. Studies of normal students and studies of students who were good decoders but poor comprehenders produced moderate effect sizes on standardized tests and large effect sizes on experimenter-developed tests. However, studies with generally poor readers (comprehension and decoding not assessed separately) resulted in null findings for standardized tests and a very large effect size on experimenter-developed tests. Thus, these students show even greater improvement than the normal readers and poor comprehenders when assessed on experimenter-developed tests but no improvement on standardized comprehension tests.

In examining the results by grade level, the researchers found evidence that RT was effective in fourth grade through college. There was insufficient evidence to make that claim for third-graders, because the three third-grade studies assessed the students with standardized tests and obtained non-significant results (Rosenhine & Meister, 1994). Because studies with older students also obtained non-significant results on standardized tests, it seems imprudent to conclude that RT is ineffective with young students. More recent studies, in fact, have provided quantitative (Boulware-Gooden, Carreker, Thornhill, & Joshi, 2007) and qualitative (Myers, 2005) evidence for the effectiveness of RT in primary grades, although the components are often modified to fit the needs of the teachers and students (Hacker & Tenet, 2002).

Pressley and colleagues (Brown, Pressley, Van Meter, & Schuder, 1996) developed a program called Transactional Strategies Instruction (TSI), which is based on instruction in connecting a text to prior knowledge, collaborative discussion emphasizing student application of strategies, and eventual internalization of the strategies that the group used to come to a consensual constructed interpretation. Like many other reading interventions involving metacognitive strategies, TSI entails long-term training with explicit instruction and modeling of the strategies as well as a focus on why the strategies are important and

when and where they are most effectively applied. Unlike many other interventions, however, TSI is focused on personal interpretations of texts.

A year-long quasi-experimental study of low-achieving second-graders (Brown et al., 1996) compared TSI instruction and more traditional instruction. TSI-trained teachers were compared with teachers who had no TSI training (matched on a variety of measures). Strategy awareness and achievement were assessed for six low-achieving students from each classroom, matched on reading comprehension scores. Students of TSI teachers reported more awareness of comprehension and word-level strategies, used more strategies on their own, and did better on literal recall of story content than the students of the non-TSI trained teachers. They also performed significantly better on both the word skills and comprehension subtests of the Stanford Achievement Test, an unusual result for studies of metacognitive reading strategies.

Collaborative Strategic Reading (CSR) is another multistrategy program designed to improve the comprehension of expository text (Vaughn, Hughes, Schumm, & Klingner, 1998). It features four strategies that students apply before, during, and after the reading process; students work in peer-mediated pairs or small groups. The strategies include: (1) previewing the text to be read, using prior knowledge of the topic; (2) monitoring comprehension and applying fix-up strategies; (3) getting the gist of each piece of text (e.g., a paragraph) by identifying the main character or object in the text and the most important information about that character or object; (4) using a wrap-up strategy to summarize key ideas and generate questions. This instructional package has been shown to improve reading comprehension in the elementary grades above fourth grade and middle school.

Vaughn et al. (2000) compared CSR instruction with a program that emphasized fluency on both fluency and comprehension outcome measures in third-graders. There were no differences between the two types of instruction. From pre-test to post-test, students assigned to both programs showed significant improvements in the rate of reading and number of correct words read per minute. Neither fluency nor comprehension instruction led to improvement in reading accuracy or in comprehension, as measured by the Gray Oral Reading Test. Results were comparable for low-achieving students and students with significant reading problems. Thus, like the rather similar RT, the effectiveness of CSR is more pronounced with children who are above the age of primary-grade students.

Self Regulated Strategy Development (SRSD), a well-researched program, focuses on teaching writing to learning disabled and at-risk students (Graham & Harris, 2003). This approach focuses on explicitly teaching both the strategies and the content necessary to write a coherent essay or paper. In addition, the program fosters self-regulation of the writing process by teaching metacognitive skills such as planning and monitoring, and motivating the students by emphasizing the importance of effort in the writing process. In a meta-analysis of studies evaluating SRSD in writing instruction (including one study of second-graders and one study of third-graders), Graham and Harris (2003) found that instruction that incorporated SRSD improved the quality, structure, and length of writing of both normally developing students and students with learning disabilities. No results on standardized tests were reported in this meta-analysis.

Integrating Listening and Reading Comprehension Instruction

Until recently school policy has been to wait until children have a fairly substantial level of decoding ability, sometimes to the point of fluency, before starting comprehension instruction. Recently, however, researchers have been interested in the possibility that early comprehension instruction could be productively initiated before students acquire

decoding skill. This has prompted an interest among reading researchers in the literature on oral language comprehension.

Kendeou, van den Broek, White, and Lynch (2007) conducted a longitudinal study in which they first looked at the oral language comprehension skills of 4-year-old children. The children were presented with aural and televised stories and demonstrated their comprehension through recall and by answering questions tapping factual and inferential knowledge of story events. At ages 8 and 10 these measurements were repeated, and the children were also given written stories. The comprehension measures from all three media were highly interrelated. Moreover, the comprehension scores at age 4 predicted later reading comprehension scores, indicating that the comprehension skills that are developed in pre-school contribute to students' later reading skills. The authors also found that these early comprehension skills develop separately from basic language skills (e.g., decoding) and vocabulary. These findings provide a good first step in making a case for early comprehension instruction using listening (storytelling), as well as other non-textual media such as television, before children start to read.

Williams and her students (Williams, Brown, Silverstein, & deCani, 1994) developed the Theme Scheme, a narrative text comprehension program whose goal is to improve comprehension as manifested in both oral and written language. It teaches students to go beyond the plot level to the identification of the story's theme. It consists of a series of questions designed to help organize the important story components and derive the thematic material and ends with questions that help extend the theme to specific and often personal scenarios. The first four questions focus on the important plot components from which a theme concept can be derived (main character, problem solution, and outcome). The next four questions encourage students to make the judgments that, when combined with the theme concept, lead to the identification of the theme. They are: (1) Was what happened good or bad? (2) Why was it good or bad? (3) The main character learned that he/she should . . . (4) We should . . . The final two questions are: (1) When is it important to . . .? (2) In what situation is it easy (or difficult) to . . .? While this Theme Scheme program represents a direct and structured approach to instruction, it also incorporates a substantial amount of class discussion that emphasizes metacognitive elements such as reflection and self-monitoring.

The program was evaluated in second- and third-grade classrooms in high-poverty schools (Williams et al., 2002). Compared with more traditional instruction, the program led to improved comprehension and identification of previously studied themes. The findings indicated that at-risk children at all achievement levels, including those with disabilities, were able to achieve some degree of abstract, higher-order comprehension when given instruction that combined structured lessons, a strategy (the Theme Scheme), and discussion. However, the program did not help students apply a theme to real-life situations or identify and apply non-instructional themes. The program had been successful in these respects when used with middle-school students, including those with learning disabilities. Perhaps this discrepancy in findings is an indication of the difficulty that abstract thinking poses for younger children.

A later version of the program contained field-tested activities that follow the three-step process for writing in the SRSD model (Graham, Harris, & Troia, 2000). The three steps are (1) writing answers to the Theme Scheme questions; (2) goal setting, brainstorming and organizing (the students compose an ending to a story according to a particular theme); (3) self-monitoring (the students use the Theme Scheme to evaluate whether they have fulfilled their goal for the ending they composed). This sequence of activities serves as the foundation for writing instruction within the Theme Scheme program.

Williams and her students have also developed instructional modules that will be

combined into a year-long second-grade program teaching expository text comprehension (Close Analysis of Texts with Structure: CATS; Williams, Hall & Lauer, 2004). The modules integrate listening and reading comprehension, embedding explicit instruction about text structure in social studies or science content. They follow an instructional model that includes explanation and modeling by the teacher, guided practice, and independent practice. The modules teach strategies that are used to comprehend text structures commonly found in texts: sequence, compare/contrast, and cause/effect. The text structures are introduced through lessons using familiar content. Then the students analyze short, well-structured social studies or science paragraphs using clue words, graphic organizers, and strategy questions.

Individual CATS modules have been evaluated successfully in at-risk schools. One module taught the compare/contrast text structure embedded in content instruction about animal classification. An evaluation of this module (Williams et al., 2005) randomly assigned teachers to one of three experimental conditions: (1) the text structure program; (2) a content program that used the same materials, including the well-structured target paragraphs, but did not focus on compare/contrast structure; and (3) a no-instruction control. The text structure program improved students' ability to comprehend compare/contrast texts. They demonstrated transfer to non-instructed compare/contrast texts. Moreover, the text structure instruction did not detract from the students' ability to learn new content. A similar program that focused on cause/effect structure through social studies content was evaluated according to the same experimental design and showed the same pattern of results (Williams et al., 2007). These findings provide evidence that formal, explicit instruction in expository text comprehension is feasible and effective as early as the second grade, an outcome that depended on the integration of listening and reading comprehension activities.

The instruction seeks to reduce the burden of executive functioning by internalizing and automatizing the necessary cognitive and metacognitive processes. The goals of the strategy instruction portion of the program are both to increase the number of comprehension strategies available to the children and to provide practice using these strategies so that the students will be able to use them on their own. There is a focus not only on what strategies are used for a certain passage but also why these strategies are effective for that passage.

The students in our programs have many opportunities to practice and enhance their metacognition. As one example, the program teaches them to identify the text structure through the recognition of clue words and a main idea sentence. After they can identify text structures, they learn to match strategies to the appropriate structure, and they decide which set of generic questions and which graphic organizer will aid their comprehension. Moreover, they are able to explain the basis on which they matched the strategies with the text structure. The ability to use these explicitly taught strategies to identify a text structure and then reflect on the strategies that were used is evidence of metacognition in these young students. As another example, the students monitor their comprehension when they participate in class discussions and integrate other students' contributions with what the teacher has to say about them.

There is one aspect of metacognitive instruction that we do not incorporate into our program, however. We do not ask the students to reflect on the nature of strategies in general or why we use strategies (Williams, 2002). We do not call attention to their mental state or ask them to think aloud about their cognitive processes. Duffy et al. (1987) and Pressley (2005) have suggested that this type of instruction can be beneficial for young students, and there are many ways to insert this type of activity, e.g., by encouraging discussion of students' individualized learning styles (Schreiber, 2005).

Others, however, feel that this type of instruction may be inappropriate at such an early age. Clay (1998), for example, suggested that it is not necessary for children to be aware of and be able to explain the processes that are going on inside their heads while they are reading, and that asking them to do so can interfere with fluent reading. She argued that this awareness develops without explicit attention being given to it, as children gain more experience with reading, as long as they have proper teacher support. Chall and Squire (1991) and McKeown and Beck (this volume) hold similar views. At this point, when the literature contains insufficient evidence one way or the other to make an informed decision, we hold to the latter view. We agree with McKeown and Beck (this volume) that attention to one's own mental processes may distract students from placing their full attention on the text and on the message that the text conveys.

In sum, the research suggests that children in the primary grades can be taught to use metacognitive strategies to improve reading comprehension and, to a lesser extent, writing proficiency. Instructional programs that incorporate explicit metacognitive strategy instruction have been found to increase students' knowledge of strategies and performance on experimenter-constructed tests, but they rarely result in gains beyond traditional instruction on standardized comprehension tests. In order to facilitate the introduction of metacognitive strategies to these young students, recent programs have sought to integrate the strategies into listening as well as reading comprehension activities. The goal of these programs is to expand the repertoire of strategies that children use to comprehend passages and to provide explicit instruction on when and how to use the strategies. This instruction appears to be effective in improving the comprehension of young students. The degree to which these young students need to be metacognitive in their implementation of the strategies is an important area for future research.

Teaching Metacognition to Teachers

Despite the great amount of enthusiasm concerning the effectiveness of comprehension strategies, there is not a great deal of attention given to comprehension strategy instruction in actual classrooms. We do know from classroom observations conducted by Moely et al. (1992) that strategy instruction, when it does occur, is more prevalent in second and third grade than earlier or later in elementary school.

Implementation of strategy instruction in the context of the actual classroom has proved problematic. It is difficult to communicate what is meant by "teaching strategies and not skills." Moreover, proficient reading involves much more than implementing individual strategies. It involves an ongoing adaptation of many cognitive and metacognitive processes. Teachers must be very skillful—and strategic—in their instruction. They must be able to respond flexibly to students' needs for feedback as they read. In order to do this, teachers must themselves have a firm grasp not only of the strategies they are teaching the children but also of instructional strategies that they can use to achieve their goal. This type of teaching is not easy, and teachers have not been prepared to do it.

An important study by Duffy and Roehler and their colleagues (Duffy et al., 1987) investigated the effect of training teachers to give explicit explanation of the reasoning and mental processes involved in reading. This Direct Explanation (DE) approach does not teach individual strategies, but rather focuses on helping students to (a) view reading as a problem-solving task that requires the use of strategic thinking and (b) learn to think strategically about solving reading comprehension problems. The implementation of the DE approach requires specific and intensive teacher training on how to teach the traditional reading skills found in basal readers *as strategies*. In Duffy et al.'s 1987 study of low-level third-grade readers, teachers were randomly assigned to one of two conditions.

Treatment teachers were shown how to provide explicit explanations, to analyze the skills presented in the basal reading texts, and to recast them as problem-solving strategies. There were 12 hours of training, which included one-on-one coaching, collaborative sharing among the teachers, observations of lessons and feedback, and videotaped model lessons. The comparison teachers were trained in classroom management procedures.

Student awareness of strategic reasoning was assessed in interviews throughout the year-long treatment. The students of the DE-trained teachers had higher levels of awareness of specific reading strategies and a greater awareness of the need to be strategic when reading than the comparison teachers' students.

Although there was no difference between the two experimental conditions in the ability to use the basal skills they had been taught, the students of DE-trained teachers had a greater ability to reason strategically when reading. They also reported using strategic reasoning more frequently when actually reading connected text. On a standardized test (the Stanford Achievement Test), students of the DE teachers outperformed the students of the comparison teachers; the difference was significant on the word skills subtest but not on the comprehension subtest. On a second standardized test (the Michigan Educational Assessment Program), administered five months after completion of the study, the students of the DE-trained teachers maintained their superiority.

These results indicate that training teachers to teach metacognitive reading strategies to third-graders with reading difficulties can be effective, and that gains can be seen on both experimenter-designed and standardized measures. The authors attributed the lack of difference on the comprehension subtest of the SAT to both the limited amount of time that students were exposed to the intervention and the differences between the focus of standardized comprehension tests and what strategic readers actually do while reading. It should also be noted that the instructions in many of the classrooms were oriented toward acquisition of word-level processes and not what we usually consider comprehension processes.

Classroom-based naturalistic studies like this one are not easy to do; they require substantial funding, collaboration between researchers and school personnel, and a great deal of preliminary descriptive and correlational work. It is not surprising that there are very few such studies (Williams, 2002). The small amount of data that we do have, however, indicates clearly that instructional methods that generate high levels of student involvement and require substantial cognitive and metacognitive activity during reading can have positive effects on reading comprehension. Moreover, providing teachers with instruction that helps them use these methods leads to students' awareness and use of strategies, which in turn can lead to improved reading comprehension.

We know that it is easier to demonstrate the positive effects of strategy instruction in the context of well-controlled "laboratory-like" studies than in the actual classroom. We also know that strategy instruction is not commonly seen in schools and that to implement strategy instruction successfully in real-life classrooms is very challenging. But it is possible to develop programs that focus on teaching teachers what they need to know in order to be successful comprehension instructors. Many questions can be raised. How extensive does teacher preparation have to be? Are direct explanation and/or collaborative discussion essential components of successful instruction? Is successful instruction the result of the particular strategies that have been taught in the studies reported here, or would a broader repertoire of instructional activities also be effective? Are other factors involved? In light of the findings to date, more research is certainly warranted, and a focus on teaching teachers what they need to know about comprehension strategy instruction appears to be the area within comprehension instruction that has the potential for moving the field forward (Williams, 2002).

Concluding Thoughts

The research of the last few decades has led to significant progress in our understanding of the nature of reading comprehension and how to teach it. The important role that cognitive and metacognitive strategies play has been documented extensively. But only recently has attention turned to the primary grades, and more research is needed at this level, especially within the natural setting of the classroom.

The evidence that comprehension strategy instruction can be effective for primary-level students comes mostly from studies that evaluate instructional programs that contain a variety of components and that are designed to be used as complete packages. Most of the components of these programs are justifiable in terms of theory. However, the studies have not been designed to determine which of them are in fact effective. Different experimental designs are needed to analyze these programs and get answers to this sort of question.

Teaching comprehension strategies effectively involves a high level of proficiency and flexibility on the part of the teacher. This requires substantial and intensive preparation. It is here, in our opinion, that further research should focus: on the development of effective instructional methods to help teachers understand the importance of metacognition. Only when teachers are aware of what their own comprehension entails will they be able to monitor their students' reading and provide the right instruction for them. It is even more important for teachers to be metacognitive than it is for their students. Teachers must be reflective about what it is that they are doing when they teach, so that they can better evaluate how their instruction is affecting their students.

It is not surprising that theory of mind research on very young children and the applied metacognitive research on older students have developed independently. However, given the recent surge of interest in teaching primary students strategies for reading comprehension (as well as for other academic skills), it would be useful to make efforts to bridge the gap that currently exists between the two strands of research. A start has been made. Investigators have been looking for links between theory of mind tasks and metalinguistic tasks relevant to reading. Farrar, Ashwell and Maag (2005) have shown that performance on the unexpected location false belief task is correlated with performance on phonological awareness tasks, in particular rhyming, even after controlling for verbal intelligence and age. Both the theory of mind task and the language tasks involve the ability to deal with conflicting representations of the same situation (representational flexibility): in a rhyming task, one must ignore semantic associations and attend to sound correspondences. Similarly, Doherty and Perner (1998) found a link between the same theory of mind task and children's understanding of synonyms, another language task requiring representational flexibility.

Longitudinal studies of preschool children, followed until they achieve literacy, would be of benefit, especially if a wide variety of linguistic measures were used. Such studies, which would suggest the causal direction of relationships between theory of mind skills and language (pre-reading) skills, might foster further theory development and might also help identify aspects of language that could be taught to young children as preparation for developing the appropriate metacognition necessary for successful reading comprehension instruction.

Even in the face of the great progress that has been made, and the reliance that we are currently placing on comprehension strategies, there is a caveat. We believe that the strategies that we are teaching our young students will serve them well when they are adults. But we do not really know whether the strategies that we are teaching our young students are those that proficient readers use in making sense of text. Perhaps that does not matter; it may be that the effectiveness of strategy instruction does not really depend on the

particular strategies that are taught. It may simply be that strategy instruction forces the students to pay attention and think about what they have read (Gersten, Fuchs, Williams, & Baker, 2001); whatever unspecified cognitive processes are at work is what is responsible for the improved comprehension. The greater effectiveness of instruction when it occurs in the context of small interactive groups (Swanson & Hoskyn, 1998) might be explained in the same way. Listening and responding to others in a discussion is an effective way to force attention and thoughtfulness to the topic at hand. We should not close our eyes to the possibility that approaches to comprehension instruction other than teaching strategies may be worth pursuing.

References

- Baker, L. (2002). Metacognition in comprehension instruction. In C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 77–95). New York: Guilford.
- Baron-Cohen, S. (2001). Theory of mind and autism: A review. *International Review of Research in Mental Retardation*, 23, 169–184.
- Boulware-Gooden, R., Carreker, S., Thornhill, A., & Joshi, R. M. (2007). Instruction of metacognitive strategies enhances reading comprehension and vocabulary achievement of third-grade students. *The Reading Teacher*, 61(1), 70–77.
- Brenna, B. A. (1995). The metacognitive reading strategies of five early readers. *Journal of Research in Reading*, 18(1), 53–62.
- Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453–482). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88(1), 18–37.
- Chall, J. S. & Squire, J. R. (1991). The publishing industry and textbooks. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. David Pearson (Eds.), *Handbook of Reading Research: Vol. 2* (pp. 120–146). White Plains, NY: Longman Publishing Group.
- Clay, M. M. (1998). Literacy awareness: From acts to awareness. In M. M. Clay (Ed.), *By different paths to common outcomes* (pp. 41–84). York, ME: Stenhouse Publishers.
- Cox, B. E. (1994). Young children's regulatory talk: Evidence of emergent metacognitive control over literary products and processes. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 733–757). Newark, DE: International Reading Association.
- Doherty, M. & Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development*, 13, 279–305.
- Duffy, G. G. (2005). Developing metacognitive teachers: Visioning and the expert's changing role in teacher education and professional development. In S. E. Israel, *Metacognition in literacy learning: Theory, assessment, instruction, and professional development* (pp. 299–314). Mahwah, NJ: Lawrence Erlbaum Associates.
- Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., Vavrus, L. G., Wesselman, R., Putnam, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, 22(3), 347–368.
- Farrar, M. J., Ashwell, S., & Maag, L. (2005). The emergence of phonological awareness: Connections to language and theory of mind development. *First Language*, 25, 157–172.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231–235). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906–911.
- Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development*, 24, 15–23.

- Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research, 71*(2), 279–320.
- Glaubman, R., Glaubman, H., & Ofir, L. (1997). Effects of self-directed learning, story comprehension, and self-questioning in kindergarten. *The Journal of Educational Research, 90*(6), 361–374.
- Graesser, A. C. (2007). An introduction to strategic reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 3–26). Mahwah, NJ: Lawrence Erlbaum Associates.
- Graham, S. & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 323–344). New York: Guilford.
- Graham, S., Harris, K. R., & Troia, G. A. (2000). Self-regulated strategy development revisited: Teaching writing strategies to struggling writers. *Topics in Language Disorders, 20*, 1–14.
- Griffith, P. & Ruan, J. (2005). What is metacognition and what should be its role in literacy instruction? In S. Israel & C. Block (Eds.), *Metacognition in literacy learning: Theory, assessment, instruction, and professional development* (pp. 3–18). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hacker, D. J. & Tenen, A. (2002). Implementing reciprocal teaching in the classroom: Overcoming obstacles and making modifications. *Journal of Educational Psychology, 94*(4), 699–718.
- Haines, D. J. & Torgessen, J. K. (1979). The effects of incentives on rehearsal and short-term memory in children with reading problems. *Learning Disability Quarterly, 2*, 48–55.
- Juliebö, M., Malicky, G., & Norman, C. (1998). Metacognition of young readers in an early intervention programme. *Journal of Research in Reading, 21*(1), 24–35.
- Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. (2007). Comprehension in preschool and early elementary children: Skill development and strategy interventions. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 27–45). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kinnunen, R., Vauras, M., & Niemi, P. (1998). Comprehension monitoring in beginning readers. *Scientific Studies of Reading, 2*(4), 353–375.
- Kreutzer, M. A., Leonard, C., & Flavell, J. H. (1975). An interview study of children's knowledge about memory. *Monographs of the Society for Research in Child Development, 40*, 1–60.
- Kuhn, D. (2000). Theory of mind, metacognition, and reasoning: A life-span perspective. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 301–326). Hove: Psychology Press.
- Lockl, K. & Schneider, W. (2006). Precursors of metamemory in young children: The role of theory of mind and metacognitive vocabulary. *Metacognition and Learning, 1*, 15–31.
- Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. *Child Development, 48*(3), 986–992.
- Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development, 50*(3), 643–655.
- Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622–646.
- Moely, B. E., Hart, S. S., Leal, L., Santulli, K. A., Rao, N., Johnson, T., Hamilton, L. B. (1992). The teacher's role in facilitating memory and study strategy development in the elementary school classroom. *Child Development, 63*, 653–672.
- Myers, P. A. (2005). The Princess Storyteller, Clara Clarifier, Quincy Questioner, and the Wizard: Reciprocal teaching adapted for kindergarten students. *The Reading Teacher, 59*(4), 314–324.
- National Reading Panel (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC: National Institute of Child and Human Development.
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 2*, 117–175.
- Paris, S. G. & Flukes, J. (2005). Assessing Children's Metacognition About Strategic Reading. In

- S. E. Israel, C. Collins Block, K. L. Bauserman, & K. Kinnucan-Welsch (Eds.), *Metacognition in Literacy Learning* (pp. 121–140). Mahwah, NJ: Lawrence Erlbaum Associates.
- Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology, 76*(6), 1239–1252.
- Paris, S. G. & Jacobs, J. E. (1984). The benefits of informed instruction for children's reading awareness and comprehension skills. *Child Development, 55*, 2084–2093.
- Pressley, M. (2005). Final reflections: Metacognition in literacy learning: Then, now, and in the future. In S. E. Israel, *Metacognition in Literacy Learning: Theory, Assessment, Instruction, and Professional Development* (pp. 391–411). Mahwah, NJ: Lawrence Erlbaum Associates.
- RAND Reading Study Group. (2002). *Reading for Understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: Rand.
- Rosenshine, B. & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*(4), 479–530.
- Ruan, J. (2004). Bilingual Chinese/English first-graders developing metacognition about writing. *Literacy, 38*(2), 106–112.
- Samuels, S. J. (2002). Reading fluency: Its development and assessment. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 166–183). Newark, DE: International Reading Association.
- Schreiber, F. J. (2005). Metacognition and self-regulation in literacy. In S. E. Israel, C. C. Block, K. L. Bauserman, & K. Kinnucan-Welsch (Eds.), *Metacognition in Literacy Learning: Theory, Assessment, Instruction, and Professional Development*, pp. 215–239. Mahwah, NJ: Lawrence Erlbaum Associates.
- Slade, L. & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology, 23*, 117–141.
- Swanson, H. L. & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research, 68*(3), 277–321.
- Vaughn, S., Chard, D. J., Bryant, D. P., Coleman, M., Tyler, B.-J., Linan-Thompson, S., & Kouzekanani, K. (2000). Fluency and comprehension interventions for third-grade students. *Remedial and Special Education, 21*(6), 325–335.
- Vaughn, S., Hughes, M. T., Schumm, J. S., & Klingner, J. K. (1998). A collaborative effort to enhance reading and writing instruction in inclusion classrooms. *Learning Disability Quarterly, 21*, 57–74.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684.
- Williams, J. P. (2002). Reading comprehension strategies and teacher preparation. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 243–260). Newark, DE: International Reading Association.
- Williams, J. P., Brown, L. G., Silverstein, A. K., & deCani, J. S. (1994). An instructional program for adolescents with learning disabilities in the comprehension of narrative themes. *Learning Disability Quarterly, 17*, 205–221.
- Williams, J. P., Hall, K. M., & Lauer, K. D. (2004). Teaching expository text structure to young at-risk learners: Building the basics of comprehension instruction. *Exceptionality, 12*(3), 129–144.
- Williams, J. P., Hall, K. M., Lauer, K. D., Stafford, K. B., DeSisto, L. A., & deCani, J. S. (2005). Expository text comprehension in the primary grade classroom. *Journal of Educational Psychology, 97*(4), 538–550.
- Williams, J. P., Lauer, K. D., Hall, K. M., Lord, K. M., Gagga, S. S., Bak, S.-J., Jacobs, P. R., & deCani, J. S. (2002). Teaching elementary school students to identify story themes. *Journal of Educational Psychology, 94*, 235–248.
- Williams, J. P., Nubla-Kung, A. M., Pollini, S., Stafford, K. B., Garcia, A., & Snyder, A. E. (2007). Teaching cause-effect text structure through social studies content to at-risk second graders. *Journal of Learning Disabilities, 40*(2), 111–120.

Part II

Metacognitive Strategies

4 Question Generation and Anomaly Detection in Texts

José Otero

Universidad de Alcalá

Questioning is an important metacognitive resource, as was recognized nearly 25 years ago by Gavelek and Raphael (1985): “. . . question asking represents one of the primary means by which individuals are able to foster their own comprehension and as such represents a powerful metacognitive activity” (p. 114). Readers who attempt to understand a text—the primary form of learning under focus in this chapter—may generate questions that they consider necessary for comprehension. The questioning activity involves monitoring, controlling, and specifically regulating comprehension by looking for information relevant to a reader’s comprehension goal.

Question asking and answering in various contexts have been studied in artificial intelligence, communication, education, linguistics, philosophy, psychology, and sociology. Early studies conducted by developmental psychologists at the beginning of the 20th century, starting with the influential work of Piaget (1955) on children’s questions, were predominantly descriptive. Researchers categorized questions according to variables, such as form, function, or source (Davis, 1932), or examined the functions and other characteristics of questions according to stages of questioning development (Lewis, 1938). A plethora of studies followed, including those that examined questioning by humans from various viewpoints (see Graesser, McMahan, & Johnson [1994] for a review) and the more current investigations of question asking and answering by machines (Light, Mann, Riloff, & Breck, 2001; Rus, Cai, & Graesser, 2007).

Different types of questions are considered in these studies. Graesser et al. (1994) distinguish four kinds of questions asked in naturalistic conversation: questions asked to (a) address knowledge deficits, (b) monitor common ground, (c) coordinate social action, and (d) control conversation and attention. Although all of them play important roles, this chapter considers only the first type of questions. Information-seeking questions address knowledge deficits by requesting information from particular sources. These questions have been singled out by many authors as being prototypical, sincere, or genuine questions (Berlyne & Frommer, 1966; Flammer, 1981; Otero & Graesser, 2001; Ram, 1991; Van der Meij, 1994). This chapter focuses on the generation of information-seeking questions when subjects process textual information. Question generation is conceived as an attempt to surmount obstacles found when questioners try to achieve particular goals. The chapter first provides a brief review of question generation mechanisms. Then there is an analysis of three basic elements of question generation processes: obstacles, goals, and knowledge of the questioner.

Question Generation Mechanisms

A fundamental challenge in question asking research is to specify the components of the question generation mechanism. An explicit acknowledgment of the importance of

understanding these mechanisms is not new (Graesser, Person, & Huber, 1992), but the systematic study of the question generation mechanisms has received less attention than other problems in question asking research, such as the strategies to develop inquisitive readers (Graesser, McNamara, & VanLehn, 2005; King, 1994) and the effects of questions on comprehension (Craig, Sullins, Witherspoon, & Gholson, 2006; Rosenshine, Meister, & Chapman, 1996; Taboada & Guthrie, 2006).

According to Dillon's (1990) componential analysis of student questioning, the first stage of asking an information-seeking question consists of an initial state of perplexity. This is followed by formulation and expression of the question in words. Graesser and McMahan (1993) identify similar stages in formulating questions to correct knowledge deficits: anomaly detection, question articulation, and social editing. All of these stages are essential in the overt formulation of a question. For instance, the importance of social barriers in student question asking has been substantiated in studies conducted by Karabenick (1996) and Van der Meij (1988). However, this review specifically focuses on the first stage: the process of anomaly detection leading to question asking.

A *knowledge deficit* hypothesis and a *knowledge clash* hypothesis have been considered as alternative mechanisms to recognize anomalies leading to questioning (Otero & Graesser, 2001). According to the former, questioning is conceived as being driven by a lack of knowledge. For example, readers who notice gaps in their lexical knowledge ask questions on the meaning of unknown words. In fact, poor readers heavily rely on the "lexical standard" for comprehension monitoring (Baker, 1985). However, question generation cannot be explained by knowledge gaps alone. In a widely cited article, Miyake and Norman (1979) ruled out a simple identification of lack of knowledge as source of questioning: "... any theory of question asking in learning cannot simply use the gaps in a person's knowledge as the source of questions" (p. 364). In their study, novices reading a difficult manual asked *fewer* questions than did experts. This finding opens up the possibility of an alternative mechanism for question generation: a knowledge clash hypothesis. According to this hypothesis, questioning depends on the inconsistencies or incompatibilities between external information and a questioner's pre-existing knowledge. This hypothesis leads to the prediction that readers with more knowledge will ask more questions because there is a greater probability of inconsistencies between knowledge and the external information.

A reconciliation of the two hypotheses can be achieved by formulating an *obstacle+goal* hypothesis. According to this hypothesis, question generation is driven by a recognition of *obstacles* when attempting to attain a particular *goal*. Obstacles to goals occur when someone tries to achieve a practical objective by performing a procedure, but also when someone tries to understand a piece of written information. A traveller encounters an obstacle when there is a knowledge deficit about appropriate public transportation, but wants to achieve the practical objective of getting to a museum. The traveller may ask someone the question "How do I get to the museum?" in an attempt to overcome the obstacle. But an obstacle also is found by a reader who attempts to understand a text. Consider the science student quoted in the study of Costa, Caldeira, Gallástegui and Otero (2000) who thinks that clouds are made of water vapor. The student encounters an obstacle when she attempts to achieve the goal of building an adequate mental representation of clouds as described in the information provided: "I was always told that clouds are water vapor . . . why then do they say that these [clouds] are not water vapor?" (p. 610). An obstacle is found (clouds are not made of water vapor) in attaining an appropriate mental representation of the situation described in the text. The student's knowledge about the composition of clouds clashes with the information provided, and triggers a question.

A number of researchers have proposed that question generation is driven by the obstacles encountered when trying to reach a goal. Flammer (1981) conceived questions as “means used by the questioner to obtain specific information which should help him or her to plan appropriate actions towards certain goals” (p. 409). Berlyne and Frommer (1966), working within a behavioural tradition, conceived of questioning as behaviour directed toward the acquisition of knowledge, motivated by a condition of high drive induced by conceptual conflict—an obstacle. The conflict may be due to properties of external stimuli, such as complexity or incongruity. A view of questioning being driven by obstacles is also consistent with Lewis’s (1938) idea that one of the chief characteristics of questions is that “the questioner does not proceed further with his activity until he receives a reply which determines this activity” (p. 151). Graesser and McMahan (1993) explicitly investigated whether anomalous information caused an increase in questions generated by college students. This “anomaly hypothesis” was contrasted with an “obstacle hypothesis” according to which only certain anomalies were considered to be real obstacles in the proposed reading tasks. Ram (1991) regarded questions on stories as arising “from an interaction between the interests and goals of the understander and the information provided by the environment” (p. 275). Obstacles and goals driving questioning have been considered also by Schank (1986a, 1986b) in the context of his work on explanation and understanding. According to Schank (1986a), explanations may be seen as attempts to answer questions generated by an understanding system. These questions are caused by situations that do not correspond to expectations, that is, situations that are anomalous in some way. However, although people’s understanding is less than perfect, not everything is felt to be anomalous or questionable: “What actually needs to be explained? . . . The answer, naturally, must involve *goals*” (Schank, 1986b, p. 25; emphasis in original).

Information-seeking questions asked on texts are also triggered by obstacles in reaching goals. A reader’s immediate goal while comprehending text is to create an internal representation of the discourse that is appropriate for the required task. Obstacles may be encountered in this attempt (Yuill & Oakhill, 1991) and questions may be asked to overcome them. For instance, a question may be asked by a reader who finds an unknown word, a situation addressed by the knowledge deficit hypothesis. A question such as “What does X mean?” may be understood as a request for information to surmount a lexical obstacle. It blocks the reader in the attainment of the tacit goal of creating a discourse representation at the relatively simple *textbase* level, that is a representation of a text’s meaning in terms of a propositional network. Of course, readers frequently attempt mental representations of discourse beyond the textbase level, and obstacles may also appear in that cognitive endeavor. For example, questions may be asked when a reader does not know the temporal or spatial circumstances of a described event (When is X? Where is X?), how a certain process develops (How X?) or the cause of a certain phenomenon (Why X?). All of these questions are caused by knowledge deficits or knowledge clashes, and can be regarded as driven by the recognition of obstacles to create a representation at the *situation model* level. A situation model is not a representation of the explicit text per se, but a representation of the situation that the text is about and incorporates aspects of the reader’s world knowledge.

Questions during text comprehension are therefore triggered by the obstacles found by readers who attempt to create an internal representation of discourse. Given the importance of obstacles during comprehension, it is important to identify the different possible question-generating obstacles.

Obstacles and Question Generation

At a global level, particular characteristics of stimuli have been found to create obstacles or anomalies: novelty, surprisingness, or incongruity. Berlyne and Frommer (1966) manipulated these variables and indeed found the expected effect on questioning. At a more detailed level, however, what is the nature of obstacles that produce questions about texts? Graesser et al. (1992, 1994) proposed five generation mechanisms for information-seeking questions: obstacle in plan or problem solving, deciding about alternatives that are equally attractive, gap in knowledge that is needed for comprehension, glitch in explanation of an event, and contradiction. Most of these mechanisms are relevant in question generation on texts. Contradictions that result from disagreements between the text and the reader's knowledge are relatively obvious obstacles that generate questions according to the knowledge clash hypothesis. Many studies have been conducted on the conditions under which this kind of obstacle is detected (Baker, 1985; Markman, 1979; Markman & Gorin, 1981) and on the regulatory behavior of participants once such obstacles have been found (Chinn & Brewer, 1993; Hacker, 1998; Otero, 2002; Van Oostendorp, Otero, & Campanario, 2002). However, as pointed out above, readers also ask questions that are better explained by the knowledge deficit hypothesis. For instance, they may ask the typical journalist questions (*Who, What, Where, When, Why* and *How*) on non-contradictory, consistent information, when they feel that their comprehension of a story is inappropriate, or when an incomplete explanation of an event is found. This poses the problem of clarifying what may constitute, in principle, a question triggering obstacle given (a) a potential questioner, (b) particular textual information input, and (c) a questioning context.

Two Main Levels of Obstacles When Reading Texts

Studies on question asking and comprehension monitoring suggest a first distinction between the obstacles underlying questioning. Gavelek and Raphael (1985) point out that “problems in [text] comprehension monitoring may be the result of the way in which the text was written . . . , gaps in the readers' background knowledge . . . , or some combination of the two” (p. 111). In a similar vein, Scardamalia and Bereiter (1992) distinguish between “text-based” and “knowledge-based” questions, and corresponding obstacles. The former are questions “prompted by a text and are generally about the text” (p. 178) while the latter “spring[s] from a deep interest of the child or arise[s] from an effort to make sense of the world” (p. 178). Scardamalia and Bereiter (1992) relate the two types of questions to the two levels of mental representation of discourse: the textbase and the situation model. Yuill and Oakhill (1991) investigated children's text-based comprehension problems at the level of individual words and sentences, versus discourse as a whole. The focus is understandably placed on problems caused by texts as linguistic representations, and less on problems about understanding the situation the text is about (although there may be connections between both).

Otero and Graesser's (2001) PREG model involves a closely related view of obstacles and questions asked when processing discourse. The model focuses on the mechanisms that trigger information-seeking questions when processing expository texts. The model accounts for the questions asked about the different levels of discourse representation attempted by readers. Readers having difficulties in creating a textbase representation would generate questions about the relationship between surface structure and textbase. These would be, for example, questions on meanings of words and sentences, or questions necessary to identify local relations between text propositions. Readers who try to build a

situation model may find obstacles of a different kind. They address the referents of words and sentences in the text, or other entities (objects, processes) not mentioned in the text but retrieved from a reader's knowledge base. Thus, a central assumption of the PREG model is that these two attempted levels of discourse representation constrain the obstacles detected and the information-seeking questions asked.

In summary, these studies suggest a basic distinction when considering the obstacles that motivate information-seeking questions on texts: obstacles directly related to the text as a linguistic external representation (called *text-based obstacles*) and obstacles related to the representation of the situation described by the text (called *knowledge-based obstacles*). Text-based questions arise as a response to obstacles of the first kind, while knowledge-based questions are prompted by obstacles of the second kind. A review of studies on question asking shows underlying obstacles that correspond to these two basic categories.

Obstacles According to Question Taxonomies and Question Generation Studies

An examination of question taxonomies is an obvious approach to identifying the text-based or knowledge-based obstacles underlying information-seeking questions. However, this identification is not always easy because questions are frequently categorized according to other dimensions than those defined by the generation mechanism. For example, Graesser et al.'s (1992) taxonomy is based on semantic, conceptual and pragmatic criteria, and not on generation mechanism, because the authors explicitly consider this to be a different dimension for question categorization. In many other cases, especially in the area of educational studies, question taxonomies are developed inductively without an explicit foundation. Therefore, the question categories found in many taxonomies correspond to an underlying mixture of obstacles responsible for their generation. For instance, Piaget's (1955) classical study on questioning analyzes the knowledge-based questions on everyday activities asked by a 6-year-old child to his tutor. Piaget distinguishes between *Why* questions, on the one hand, given their special importance in children's questioning, and questions not formulated as *Why*, on the other. *Why* questions may be of three types: (a) causal explanation questions, when subjects look for a cause of material events, be it an efficient or a final cause, (b) motivation questions, when a reason or motive of an action or a psychological state is sought in terms of "For . . .", and (c) justification questions, when a reason for a rule or definition is searched, including logical justifications. Questions not formulated in terms of *Why* include "reality and history" questions that are addressed to know characteristics and circumstances of facts or events, questions asked on human actions and intentions (that may overlap with previous categories), questions asked on rules and customs, and questions asked on classifications and evaluations. Although there is not a one-to-one correspondence between these question categories and a well-defined set of obstacles, Piaget's taxonomy clearly implies that not knowing efficient or final causes, or not knowing characteristics and circumstances of objects and events, are potential knowledge-based obstacles for children's understanding.

Some studies examining question asking on narrative and expository texts do explicitly consider the obstacles that are at the root of questioning. The basic rationale to ground questions on obstacles consists in linking the former to difficulties in achieving understanding. In this view, the obstacles found and the questions asked critically depend on how understanding is conceived.

Ram's (1991) work exemplifies this approach. Questions asked on stories are conceived as directly related to understanding tasks. These may be parser-level tasks, such as

pronoun resolution, or higher-level tasks such as integrating facts with what the reader already knows (Ram, 1991, p. 282). These understanding tasks result in questions necessary to perform the task. In this view, the obstacles that trigger questions when readers process stories result from an analysis of the understanding tasks and knowledge goals involved in story comprehension. Ram (1991) distinguishes four types of knowledge goals, or questions, that correspond to four types of obstacles. First, there are obstacles associated with text goals that “arise from basic syntactic and semantic analysis that need to be performed on the input text” (Ram, 1991, p. 304). They produce “text questions.” Second, there are obstacles that arise in the process of noticing similarities, matching incoming concepts to stereotypes in memory, or forming generalizations. These result in “memory questions.” Third, there are obstacles that emerge when trying to resolve anomalies or building motivational or causal explanations for events and they lead to “explanation questions.” Finally, there are obstacles found when a reader tries to identify aspects of the situation that are personally interesting or relevant to one’s own general goals. They result in “relevance questions.” The first type of questions corresponds to text-based obstacles, while the others are caused by knowledge-based obstacles.

Trabasso, Van den Broek and Liu’s (1988) model of question generation to assess comprehension of narratives is based on the same rationale: “Knowing a priori a logical possible interpretation of the text allows the specification of what the comprehender should know, which content to query, and which questions to ask. One purpose of this paper is to illustrate how a representational theory guides the generation of questions on stories” (Trabasso et al., 1988, pp. 25–26). The “logical possible interpretation” of the narrative is conceived as a general recursive transition network where causal relations play an essential role. The comprehender’s questions are constrained by this representation, both in the queried relations and queried content. Questioning rules establish that questions should be asked on causes and consequences, and should focus on the basic episodic units of goals, actions and outcomes. Therefore, adequately representing causes and consequences are conceived as the main obstacles that trigger question asking for narratives. Other obstacles may conceivably exist, for example adequately representing characters in the narrative, but the model considers them to have lower relevance.

Ishiwa, Macías, Maturano and Otero (2009) essentially used the same foundation to analyze question generation and the obstacles driving questioning when subjects tried to understand expository, scientific texts. The rationale for identifying and categorizing obstacles and questions was also based on an analysis of the goal of a reader who reads for understanding: to create an appropriate situation model representation. This has been considered by some researchers as essential for comprehension (Graesser, Singer, & Trabasso, 1994; Kintsch, 1998). Building a situation model involves generating inferences that elaborate the textbase. Consequently, the obstacles found when building a situation model are directly linked to difficulties in generating these inferences. Three classes of obstacles were distinguished and aligned to the three broad classes of inferences made in conscious understanding: associative, explanatory, and predictive (Trabasso & Magliano, 1996). Associations provide information about descriptive detail of the entities mentioned in the text. Explanations provide reasons for why something occurs. Predictions include consequences of actions or events and anticipate future occurrences. In the case of associations, readers may find obstacles in representing the objects or processes under consideration, as well as their properties, and may ask questions about them. These questions roughly correspond to Piaget’s (1955) “reality and history” questions, or to the basic journalist questions *Who, What, Where, When, and How*. The second type of obstacles consists of difficulties in justifying or explaining these entities. They result in *Why* questions addressed to know efficient causes, and final causes or human goals. Finally, there are

obstacles to foresee consequences of the represented entities, and they correspond to *What Happens Next* or *What If* questions.

Schank's (1986a, 1986b) work on explanation explicitly focuses on the nature of knowledge-based obstacles that are responsible for question generation processes. When human actions are considered, anomalies leading to questioning result from violations of a standard comprising *patterns, consequences, and reasons*: "... as observers of actions we are satisfied when the action we observe fits into a known pattern, has known consequences we can determine be beneficial to the actor, or is part of an overall plan or view of the world that we can ascribe to the author" (Schank, 1986a, p. 146). Accordingly, to understand human actions, there may be a class of obstacles and questions about the pattern into which the action may fit, on the unknown consequences that the action may have, and on the unknown reasons or goals that the actor has. This classification is similar to the one used by Ishiwa et al. (2009) for obstacles in understanding physical situations.

Graesser and McMahan (1993) explicitly tested the assumption of a relation between anomalies and questioning by deliberately introducing four types of anomalies in mathematical word problems and in stories. The anomalies introduced in the regular versions of word problems and stories were created by *deleting* a critical piece of information to solve the problem or to understand the story, by introducing a sentence *contradicting* an earlier statement, or by introducing salient or, alternatively, subtle *irrelevancies*. However, only contradictions and deletions were considered obstacles. Irrelevancies were considered anomalies that do not impede solving a problem or understanding a story. The results showed that the deletion condition consistently triggered more questions than the original versions, whereas the contradiction and salient irrelevancy only sporadically increased question asking.

In conclusion, obstacles defining information-seeking questions on texts are associated with the internal representation that the comprehender attempts to build. The specificity of the reading goals is an extremely important constraint on the question asking process. In tasks such as well-defined word problems, where a clear description of a final state exists (VanLehn, 1989), goals are well defined and obstacles may be well defined too, making questions predictable. In contrast, goals are less well defined in other situations, such as those corresponding to many school activities (Bereiter & Scardamalia, 1989). For instance, goals are less well defined, and obstacles are harder to identify, in reading tasks such as the comprehension of a story (Graesser & McMahan, 1993) or the comprehension of a scientific text that describes natural systems (Costa, Caldeira, Gallástegui, & Otero, 2000; Taboada & Guthrie, 2006). Predicting questions that may be asked in the latter situations depends on the readers' conceptualization of "comprehension." As shown by the studies reviewed above, only a few steps have been taken to identify question triggering obstacles in these situations, particularly those involving knowledge deficits. For instance, it is not easy to predict in a principled way the knowledge deficits that a science or a history student may notice when reading a textbook chapter for understanding. Therefore, predicting the questions that may be asked when tasks and goals are ill-defined turns out to be much more difficult than for well-defined tasks.

Reading Goals and Question Generation

Readers attempt different internal representations of discourse, that is, they generate different kinds of inferences, depending on reading situation and reading purpose. For instance, reading an expository text with a study goal results in more explanatory and predictive inferences than reading the same text with an entertainment goal (Van den Broek, Lorch, Linderholm, & Gustafson, 2001). Therefore, given that different goals

affect the kind of representation attempted, they should also influence the obstacles found and questions asked. In fact, the role of goals in question generation has been explicitly recognized by several researchers, most explicitly in Ram's (1991) model of question asking. According to this model, questions are equivalent to subgoals in the understanding tasks of reasoners: "Understanding is a process of relating what one reads to the questions that one already has. These questions represent the knowledge goals of the understander, that is the things that the understander wants to learn" (p. 277).

Graesser, Langston and Bagget (1993) have demonstrated how different goals influence the kind of questions asked by subjects processing written information. A program designed to learn by questioning was used by students who were asked to explore a computer database on woodwind instruments. They were given instructions to ask questions when needed. Half of the subjects were told that their goal was to design a wind instrument with certain characteristics—a goal involving deep causal knowledge. The other half were told to assemble a band to play at a New Year's Eve party—a goal requiring superficial knowledge. The questions asked were quite different in the two conditions. The frequency of causal questions was high and increased over time in the "Design" condition. In contrast, the frequency of causal questions was low and constant over time in the "Assemble" condition. This latter goal prompted subjects to mainly ask questions on taxonomic-definitional knowledge. When reading with the design goal in mind, the obstacles found seemed to consist in clarifying causal relations among the entities involved. In contrast, the representation goal in the assemble condition determined a different kind of obstacle. The students basically found difficulties to represent the entities themselves and, to a much lesser extent, the causal relations among them.

A similar effect of reading goals was found by Ishiwa, Sanjosé, Higes and Otero (2006). The participants in the study read essentially identical paragraphs under two goal conditions: (a) reading in order to "understand" a paragraph describing sailing against the wind, and (b) reading in order to solve a problem related to this description (finding out the speed of the boat). The importance of causal relations was expected to be greater in the former condition compared to the latter, because the problem could be solved through a simple algorithm provided in the paragraph. Accordingly, the number of causal obstacles found, and causal questions asked, would be different too. The results supported this hypothesis: readers generated significantly more causal questions in the understanding goal condition than in the problem-solving goal condition.

In sum, different questions are indicative of the obstacles found by questioners in their attempt to reach different goals. The immediate goal of a reader, that is, the intended text representation, may be influenced by external tasks, as shown by the previous studies. But the intended text representation may be influenced by other variables also, such as reader's knowledge. Reader's knowledge, be it world knowledge or knowledge about what constitutes an appropriate mental representation of discourse, appears as an important variable in question generation also. Its role is considered in the next section.

Reader's Knowledge and Question Generation

Since the study by Miyake and Norman (1979), questioners' knowledge has been singled out by several researchers as a key personal variable in question asking (LaFrance, 1992; Van der Meij, 1990; Graesser & Olde, 2003). However, studies on the role of knowledge in question generation have provided inconsistent results. Lack of knowledge was pointed out by Miyake and Norman (1979) as an explanation of the observed paucity of questions asked by subjects reading difficult material. Novices reading a difficult manual acted as if "they did not know what they should know to know further" (Miyake & Norman, 1979,

p. 363). In this view, knowledge has a positive influence on question generation, as predicted by the knowledge clash hypothesis. However, other studies have shown an opposite effect of knowledge on questioning (Van der Meij, 1990). Several additional variables may explain these discrepancies, because the relation between knowledge and questioning has been frequently studied in complex settings (Fuhrer, 1989). An obstacle+goal hypothesis provides some hints to examine the relation between knowledge and the process of obstacle identification that leads to question generation. Lack of knowledge may affect question generation on textual input in two ways, not altogether independent. First, it may hinder the initial representation of the textual input. Second, it may limit the representation goal of the questioner. Both of these will influence the obstacles found and the questions asked.

Knowledge and Questioning: The Initial Representation

Graesser and Olde's (2003) study illustrates how lack of knowledge may limit question generation because of a poor representation of the information input (text plus graphics). Students with variable technical knowledge were provided with illustrated texts on everyday devices, such as a cylinder lock, and a description of breakdown scenarios. They were given instructions to ask questions in order to identify the components responsible for the malfunctions. In addition to an important effect of technical knowledge on the quality of questions asked, the authors found that students with high technical knowledge tended to ask more questions focusing on likely malfunctioning parts than students with less technical knowledge. The latter asked more diffuse questions, or questions on components that would not explain the breakdown. In addition, questions asked by students with high technical knowledge had more mechanistic detail and a more precise elaboration of the device. The result is consistent with a positive influence of knowledge on the identification of obstacles because of a richer representation of the described device. Technical knowledge enabled subjects to identify objects or processes (for instance, the movement of the cam) that result in obstacles to represent the malfunctioning device. Therefore, they were able to ask corresponding questions addressing the malfunction. In contrast, low knowledge subjects were frequently unable to represent the device in enough detail to identify appropriate obstacles, and asked shallow questions such as "What kind of lock is it?"

A convergent result about limitations in obstacle identification caused by a poor initial text representation was found by Caldeira, Macías, Maturano, Mendoza and Otero (2002). They analyzed the questions asked by 18-year-old students and 6-year-old students who watched a short film on sea waves. The older students were aware of obstacles involving crests of waves or sea currents, and asked questions such as "Why do waves have crests that create a kind of tunnel before breaking up?" or "Are surface sea currents similar to deep sea currents?" However, none of the questions asked by 6-year-old students mentioned crests or currents. Following a pattern that is similar to the one found for low technical knowledge students in Graesser and Olde's (2003) study, 6-year-old students asked questions addressing very general processes (to move, to float), or simple objects such as river, or water: "Why does the river move?", "Why does water move?" These differences suggest that the only obstacles leading to questioning are those that could be described in terms of the categories available to the questioner—a relatively obvious result. Of course, this is directly related to the knowledge clash hypothesis: more available categories may result in more questioning because of a greater probability of knowledge disagreements.

A different influence of knowledge on the detection of obstacles was evidenced in the study of Otero and Kintsch (1992). Using the C-I model of discourse comprehension

(Kintsch, 1998), they explained how an unbalanced activation of readers' knowledge leads to the suppression of text information that is inconsistent with this knowledge. A highly activated proposition entering into the construction phase may result in the inhibition of others that are inconsistent with it (i.e., that are connected through negative links). This may happen when a reader holds strong beliefs on a certain topic, as frequently happens with science misconceptions. A reader in this situation would not detect obvious obstacles to comprehension such as explicit contradictions—a well-documented phenomenon in the comprehension monitoring literature (Markman, 1979; Otero, 2002). Therefore, the generation of questions to overcome these obstacles would be impossible also.

Knowledge and Questioning: The Representation Goal

Differences in the attempted representation of textual input may be due to differences in readers' tasks, as has been shown in a previous section. But, given a certain task, knowledge may influence the attempted text representation also. Piaget's (1955) study on questioning illustrates this influence. According to Piaget, a conception of the world where chance plays a very limited role explains why children from approximately 3 to 7 years of age ask so many *Why* questions. In contrast to adults, who rely on chance and contingency to explain many phenomena, children believe in a highly predictable world where chance plays a lesser role. Assuming there is very limited room for chance in their mental representations of the world, many events and states that are not problematic for adults turn out to be full of obstacles for children. When Del, the child studied by Piaget, observes his tutor finding and picking up a stick, he asks "Why is this stick bigger than you?" (Piaget, 1955, p. 181). Chance plays a role in adults' mental representations, such that not every entity in the world, or entity's feature, should be causally explained. Consequently, a characteristic such as the size of a stick found while strolling, as long as it is kept within a familiar range, does not represent an obstacle in understanding and does not lead to questioning: chance accounts for the particular size of the stick. However, having a different conception of the role of chance, Del finds an obstacle in representing this particular object and asks a question that an adult would not consider: "A child is still far from allowing its share to chance in the nexus of events and tries to find a reason for everything" (Piaget, 1955, p. 183).

A related influence of knowledge on mental representations and on the recognition of obstacles may be found when comparing scientific questions and commonsense questions about the world. Scientific knowledge of the world, and scientists' thinking, is qualitatively different from everyday knowledge and laypersons' thinking (Reif & Larkin, 1991). The goal of science is to achieve maximum predictive and explanatory power based on a minimum number of premises. Scientific comprehension criteria require rigorous deductions of the observed phenomena from these basic premises. In contrast, everyday thinking is typically fragmented and does not demand precise knowledge or long inference chains from a few basic principles. People may think that they understand processes or objects when they have enough familiarity to deal with them satisfactorily for everyday purposes. Therefore, the attempted mental representations of laypersons and of scientists trying to understand the world are quite different, and the obstacles and questions asked are different too. These differences are especially salient when it comes to the representation of states in the world and the resulting questions that may be asked about them. Laypersons may cope with surrounding states creating mental representations quite appropriate for everyday activities. For example, mentally representing the sky hardly involves any obstacle and does not trigger any question from most people in everyday circumstances. The situation is quite different for someone trying to create a scientific representation. The

color of the sky should be deduced from basic premises regarding the interaction of sunlight with the atmosphere. So the blue color of the sky could conceivably represent an obstacle and trigger a question for a scientifically minded person, such as, “Why is the sky blue and not, say, white?”* In fact, generating good questions, i.e., finding good problems, is a key component of scientific research (Loehle, 1990) and of creative behavior in general (Runco, 1994).

Closing Comment

Three basic elements of the mechanism of question generation on texts have been presented above: obstacles, goals, and readers’ knowledge. This review only addressed a fraction of the complexities of anomaly detection and question generation. However, the three components examined and an obstacle+goal hypothesis allow an interpretation of several pieces of existing research on question generation. A more sophisticated knowledge of the obstacles leading to questioning, as well as of the factors that may affect their identification, contribute to understanding how we identify ignorance and discover problems. These are important cognitive processes inside schools, and also outside of them.

Author Note

This research was funded by grants of the Ministry of Education, Spain (SEJ2005-03126 and HP2006-023). I am grateful to Adolfo López Otero and Aymée Rivera for help in the preparation of the chapter.

References

- Baker, L. (1985). How do we know when we don’t understand? Standards for evaluating text comprehension. In D. L. Forrest-Pressley, G. E. Mackinnon, T. G. Waller (Eds.), *Metacognition, cognition and human performance* (pp. 155–205). New York: Academic Press.
- Bereiter, C. & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361–392). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Berlyne, D. E. & Frommer, F. D. (1966). Some determinants of the incidence and content of children’s questions. *Child Development*, 38, 177–187.
- Caldeira, M. H., Macías, A., Maturano, C., Mendoza, A., & Otero, J. (2002, April). *Incomprehension and question-asking on natural phenomena described in texts or presented in films*. Paper presented at the 2002 Annual Meeting of the American Educational Research Association, New Orleans.
- Chinn, C. & Brewer, W. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63, 1–49.
- Costa, J., Caldeira, M. H., Gallástegui, & Otero, J. (2000) An analysis of question asking on scientific texts explaining natural phenomena. *Journal of Research in Science Teaching*, 37, 602–614.
- Craig, S., Sullins, J., Witherspoon, A., Gholson, B. (2006). The deep-level-reasoning-question effect: The role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction*, 24, 565–591.
- Davis, E. A. (1932). The form and function of children’s questions. *Child Development*, 3, 57–74.

* Olbers’ paradox, described by the German astronomer H. W. Olbers in 1823, is an illustrative example of the difference in obstacles found when attempting scientific understanding vs. everyday understanding of states. According to scientific principles about light intensity and about the distribution of stars in a static, infinite universe, the night sky should appear bright and not dark. However, the darkness of the night sky is far from constituting an obstacle to laypersons’ understanding of the physical world.

- Dillon, J. T. (1990). *The practice of questioning*. New York: Routledge.
- Flammer, A. (1981). Towards a theory of question asking. *Psychological Research*, 43, 407–420.
- Fuhrer, U. (1989). Effects of prior knowledge, crowding, and congruence of subject's and other's goals on question asking in an unfamiliar setting. *Psychological Reports*, 64, 131–145.
- Gavelek, J. R. & Raphael, T. E. (1985). Metacognition, instruction, and the role of questioning activities. In D. L. Forrest-Pressley, G. E. Mackinnon, & T. G. Waller (Eds.), *Metacognition, cognition and human performance* (Vol. 2, pp. 103–136). Orlando, FL: Academic Press.
- Graesser, A. C., Langston, M. C., & Bagget, W. B. (1993). Exploring information about concepts by asking questions. In G. V. Nakamura, R. M. Taraban, & D. Medin (Eds.), *The psychology of learning and motivation: Vol 29, categorization by humans and machines* (pp. 411–436). Orlando, FL: Academic Press.
- Graesser, A. C. & McMahan, C. L. (1993). Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. *Journal of Educational Psychology*, 85, 136–151.
- Graesser, A. C., McMahan, C. L., & Johnson, B. K. (1994). Question asking and answering. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 517–538). New York: Academic Press.
- Graesser, A. C., McNamara, D. S. & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor and iSTART. *Educational Psychologist*, 40, 225–234.
- Graesser, A. C. & Olde, B. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, 95, 524–536.
- Graesser, A. C., Person, N. K., & Huber, J. D. (1992). Mechanisms that generate questions. In T. Lauer, E. Peacock, & A. C. Graesser (Eds.), *Questions and information systems* (pp. 167–187). Hillsdale, NJ: Erlbaum.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 3, 371–395.
- Hacker, D. (1998). Self-regulated comprehension during normal reading. In D. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 165–191). Hillsdale, NJ: Erlbaum.
- Ishiwa, K., Macías, A., Maturano, C., & Otero, J. (2009). Generation of information-seeking questions when reading science texts for understanding. Manuscript under revision.
- Ishiwa, K., Sanjosé, V., Higes, J. M., & Otero, J. (2006). Influence of readers' goals on information-seeking questions asked on expository texts. Paper presented at the EARLI SIG Meeting on "Metacognition," Cambridge, UK, 18–20 July.
- Karabenick, S. A. (1996). Social influences on metacognition: Effects of coler learner questioning on comprehension monitoring. *Journal of Educational Psychology*, 88, 689–703.
- King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, 31, 338–368.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- LaFrance, M. (1992). Questioning knowledge acquisition. In T. Lauer, E. Peacock, & A. C. Graesser (Eds.), *Questions and information systems* (pp. 11–28). Hillsdale, NJ: Erlbaum.
- Lewis, M. M. (1938). The beginning and early functions of questions in a child's speech. *British Journal of Educational Psychology*, 8, 150–171.
- Light, M., Mann, G., Riloff, E., & Breck, E. (2001). Analyses for elucidating current question answering technology. *Natural Language Engineering*, 7, 325–342.
- Loehle, C. (1990). A guide to increased creativity in research—inspiration or perspiration? *BioScience*, 40, 123–129.
- Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development*, 50, 643–655.
- Markman, E. M. & Gorin, L. (1981). Children's ability to adjust their standards for evaluating comprehension. *Journal of Educational Psychology*, 73, 320–325.

- Miyake, N. & Norman, D. A. (1979). To ask a question one must know enough to know what is not known. *Journal of Verbal Learning and Verbal Behavior*, 18, 357–364.
- Otero, J. (2002). Noticing and fixing difficulties while understanding science texts. In J. Otero, J. A. León, & A. Graesser (Eds.), *The psychology of science text comprehension* (pp. 281–307). Mahwah, NJ: Erlbaum.
- Otero, J. & Graesser, A. (2001). PREG: Elements of a model of question asking. *Cognition and Instruction*, 19, 143–175.
- Otero, J. & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*, 3, 229–235.
- Piaget, J. (1955). *The language and thought of the child* (M. Gabain, Trans). New York: Meridian Books. The Noonday Press. (Original work published in 1926.)
- Ram, A. (1991). A theory of questions and question asking. *The Journal of the Learning Sciences*, 1, 273–318.
- Reif, F. & Larkin, J. (1991). Cognition in scientific and everyday domains: Comparison and learning implications. *Journal of Research in Science Teaching*, 28, 733–760.
- Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, 66, 181–221.
- Runco, M. A. (1994). *Problem finding, problem solving and creativity*. Norwood, NJ: Ablex Publishing Company.
- Rus, V., Cai, Z., & Graesser, A. C. (2007). Experiments on generating questions about facts. In *Computational linguistics and intelligent text processing* (pp. 444–455). Berlin / Heidelberg: Springer.
- Scardamalia, M. & Bereiter, C. (1992). Test-based and knowledge-based questioning by children. *Cognition and Instruction*, 9, 177–199.
- Schank, R. C. (1986a). Explanation: A first pass. In J. L. Kolodner & C. K. Riesbeck (Eds.), *Experience memory and reasoning* (pp. 139–165). Hillsdale, NJ: Erlbaum.
- Schank, R. C. (1986b). *Explanation patterns. Understanding mechanically and creatively*. Hillsdale, NJ: Erlbaum.
- Taboada, A. & Guthrie, J. T. (2006). Contributions of student questioning and prior knowledge to construction of knowledge from reading information text. *Journal of Literacy Research*, 38, 1–35.
- Trabasso, T. & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes*, 21, 255–287.
- Trabasso, T., Van den Broek, P., & Liu, L. (1988). A model for generating questions that assess and promote comprehension. *Questioning Exchange*, 2, 25–38.
- Van den Broek, P., Lorch, R. F., Linderholm T., & Gustafson M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, 29, 1081–1087.
- Van der Meij, H. (1988). Constraints on question asking in classrooms. *Journal of Educational Psychology*, 80, 401–405.
- Van der Meij, H. (1990). Question asking: To know that you do not know is not enough. *Journal of Educational Psychology*, 82, 505–512.
- Van der Meij, H. (1994). Student questioning: A componential analysis. *Learning and Individual Differences*, 6, 137–16.
- VanLehn, K. (1989) Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of Cognitive Science* (pp. 527–579). Cambridge, Massachusetts: The MIT Press.
- Van Oostendorp, H., Otero, J., & Campanario, J. M. (2002). Conditions of updating during reading. In M. Louwerse & W. Van Peer (Eds.), *Thematics: Interdisciplinary Studies* (pp. 55–75). Amsterdam: John Benjamins.
- Yuill, N. & Oakhill, J. (1991). *Children's problems in text comprehension*. Cambridge: Cambridge University Press.

5 Self-Explanation and Metacognition

The Dynamics of Reading

Danielle S. McNamara
The University of Memphis

Joseph P. Magliano
Northern Illinois University

What is the relation between *metacognition*, or *metacomprehension*, and self-explanation? This is a question that has haunted our research for over a decade. One of our guiding research goals is to better understand the processes involved in reading comprehension. To that end, we have spent the bulk of our careers studying reading processes and strategies as revealed in verbal protocols. These verbal protocols are produced under instructions to “think aloud” (Magliano, Trabasso, & Graesser, 1999; Magliano & Millis, 2003; Trabasso & Magliano, 1996a, 1996b) or to self-explain while reading (McNamara, 2004). Thinking aloud is a process of reporting whatever thoughts are available to a reader while reading a text (e.g., readers report these thoughts after reading each sentence). Self-explanation is the process of explaining text to oneself either orally or in writing. Both think aloud and self-explanation are composed of natural speech, including incomplete and ungrammatical sentences. To contrast the two, think aloud is assumed to reflect unaltered processes that would occur regardless of the think aloud process, whereas self-explanation is generally assumed to modify comprehension and learning processes. The two also differ in intended audience. Whereas during a think aloud, the audience is clearly the experimenter, the audience during a self-explanation is more ambiguous. That is, the intent of self-explanation is that the reader explains challenging information to himself or herself. Thus, the audience is the reader. However, in experimental situations, there is often someone present (a person or a computer that is providing feedback); thus, the participant may regard a listener as the audience.

We have endeavored to understand the processes and the products of these tasks (e.g., McNamara, Boonthum, Levinstein, & Millis, 2007; Trabasso & Magliano, 1996a) and their relationships with comprehension and reading ability (Magliano & Millis, 2003; Magliano, Trabasso, & Gresser, 1999; McNamara, O’Reilly, Best, & Ozuru, 2006; Ozuru, Best, & McNamara, 2004). We have also developed reading comprehension assessments (Gilliam, Magliano, Millis, Levinstein, & Boonthum, 2007) and reading interventions (McNamara, 2004; McNamara, Boonthum, Levinstein, & Millis, 2004; Levinstein, Boonthum, Pillarisetti, Bell, & McNamara, 2007) based on these tasks.

Part of our research agenda involves identifying components of reading that offer diagnostic signatures of poor comprehension; thus indicating which readers may be in greater need of training, scaffolding, or better, more cohesive texts. In one line of our research, we have found that having participants overtly produce reading strategies such as self-explanation provides a sound basis for diagnosing reading comprehension skill (e.g., Magliano & Millis, 2003; Millis, Magliano, & Todaro, 2006). Also, the use of reading strategies is crucial to comprehension under challenging circumstances, such as when

reading unfamiliar text (e.g., McNamara, 2004). Therefore, participants' knowledge and use of metacognitive reading strategies *should* offer an ideal technique for diagnosing students' need for reading remediation, such as receiving instruction in self-explanation and reading strategies.

Over the course of conducting this research, we have come to what we believe to be an irrefutable conclusion: the reading strategies that readers engage in while self-explaining or thinking aloud are heavily guided by metacognition. The act of self-explanation by its very nature requires the reader to be aware of the comprehension process. The process of self-explanation brings to the surface and externalizes the comprehension process for the reader. As such, it is a process that both requires and induces metacognition. Metacognition as it relates to reading involves the reader being aware of the reading process and knowing what to do when the reader's level of comprehension is not sufficient, or does not satisfy the reader's goals and desires. We assume that more skilled readers are more likely to engage in explanatory processes while reading because they are more aware of whether they do or do not understand the text and they know whether explanation (and other strategies) will help them to attain their comprehension goals. Regardless of reading skill, when readers are prompted to self-explain, they are generally more aware of the reading process. This awareness emerges as a consequence of the very nature of self-explanation. However, readers may or may not be aware of the various types of reading strategies that will help them to better understand the text and more effectively explain it. Thus, unprompted self-explanation both requires and induces metacognitive processes. Prompted self-explanation may not require metacognition, but naturally induces it at some level.

The product of self-explanation (i.e., the self-explanation protocol) allows us as researchers to study readers' ability and propensity to use comprehension strategies. It allows us to examine the ebb and flow of comprehension processes, which dynamically change as a function of the text as well as the reader's goals and interests. Self-explanation and think aloud both provide a moving snapshot of reading and metacognitive processes, and thus allow us to better understand comprehension processes. We believe that understanding the relationship between metacognition and self-explanation is critical to achieving the ultimate goal of better understanding the comprehension processes revealed in these tasks.

Notably, one challenge that we have encountered is how to *measure* metacognition, particularly as it relates to the reading process. This chapter tells the story of how we have approached this challenge. Unfortunately, we end with few firm answers: This challenge and question has not yet been fully solved. It is clear that a good deal more research will be needed to fully understand the issues. Nonetheless, this chapter comprises what answers we have so far.

What is Self-Explanation?

In our research, we have defined self-explanation as the process of explaining text or material to oneself either orally or in writing. These explanations are generally based on information contained in the discourse context and relevant world knowledge. Self-explanation can be initiated because it is associated with an explicit comprehension goal to explain while reading (Chi, de Leeuw, Chiu, & LaVancher, 1994; Magliano, Trabasso, & Graesser, 1999; McNamara, 2004) or it can occur naturalistically, presumably as the result of a metacognitive awareness of a need to explain what is being read (Chi, Bassok, Lewis, Reimann, & Glasser, 1989) or a search for coherence (Graesser, Singer, & Trabasso, 1994). A good deal of research has shown that readers who self-explain either

spontaneously or when prompted to do so understand more from learning materials and construct better mental models of the content (Chi et al., 1989, 1994; Chi & VanLehn, 1991; Magliano et al. 1999; Trabasso & Magliano, 1996b; VanLehn, Jones, & Chi, 1992).

Self-explanation protocols can be analyzed in order to better understand the process of self-explanation, learning, problem solving, and comprehension. Such analyses require researchers to develop coding schemes that define the products and processes involved in self-explanation. While there is little disagreement on the definition of the process of self-explanation, there is more debate on what constitutes the product of the self-explanation. Some researchers consider all that is produced during the process of self-explanation to be *self-explanation* per se (e.g., Chi & Bassok, 1989; Chi et al., 1994; McNamara, 2004), whereas others consider only those utterances that reveal inferences and knowledge building processes or provide reasons for *why* an event occurs to constitute *self-explanation* (Trabasso & Magliano, 1996a, 1996b; Coté & Goldman, 1999; Wolfe & Goldman, 2005; Graesser & Clark, 1995). This disagreement is primarily attributable to differences in the focus of the task used by the researchers, rather than fundamental theoretical differences. Researchers who are examining the task or process of self-explanation generally treat the entire explanation as a self-explanation, and examine the processes and strategies that emerge and become apparent within the self-explanation (Chi & Bassok, 1989; Chi et al., 1994; McNamara, 2004). That is, a self-explanation is the product or outcome of the process of self-explanation, regardless of its quality or characteristics. The contrasting perspective is most often adopted by researchers who are collecting think-aloud protocols, as opposed to specifically directing participants to self-explain (e.g., Trabasso & Magliano, 1996a, 1996b). When thinking aloud, a participant's task is simply to verbalize the thoughts and processes that are available consciously to the participant after every sentence in a text (Trabasso & Magliano, 1996a, 1996b), after specific sentences (Magliano & Millis, 2003), or at locations selected by the participant (Pressley & Afflerbach, 1995). In a think-aloud protocol, not all of the utterance is a self-explanation, and thus parts are often distinctly identified as going beyond the material and constituting an explanation of the material.

To better understand the nature of self-explanation, consider the protocols presented in Table 5.1, generated in response to these two sentences: *A congenital disease is one with which a person is born. Most babies are born with perfect hearts, but something can go wrong for approximately 1 in 200 cases.* On the left is the utterance generated by a participant and on the right is how it might be classified. The process of self-explanation comprises the use of a variety of strategies that support the construction of meaning, some of which may not provide reasons for why events occur and may not be considered inferences. Whereas the first four examples typically would not be considered as explanations of the sentences, the last three are increasingly closer to explanations, with the sixth being closest to what might be labeled as a *knowledge building* explanation (Coté & Goldman, 1999; Magliano, Wiemer-Hastings, Millis, Munoz, & McNamara, 2002). However, there is greater disagreement concerning what is and what is not an *explanation* than one may expect. For example, some researchers may argue that the fourth example (i.e., the bridging inference) is not an explanation because it does not go beyond the text, whereas others may argue that it is an explanation because it explains the relationship between the two sentences (Trabasso & Magliano, 1996a, 1996b). While there could be some debate on whether a statement does or does not constitute an explanation, there is usually less disagreement on the relationship of the statement to the text. For example, most researchers would agree that the third example is a paraphrase of the original text because it does not go beyond the text or link the two sentences. Likewise, most would agree that the fifth

Table 5.1 Statements (Protocols) Generated in Response To: *A congenital disease is one with which a person is born. Most babies are born with perfect hearts, but something can go wrong for approximately 1 in 200 cases*

<i>Statement</i>	<i>Classification</i>
1) Babies are cute and cuddly but cry a lot.	Nonsense/Irrelevant
2) I don't understand this.	Comprehension monitoring
3) Congenital means that you're born with it. Most babies are born with ok hearts, but sometimes something goes wrong, and this can happen to 1 out of 200 babies.	Paraphrase
4) My cousin had a problem with his heart and so I guess it was congenital.	Elaboration
5) So, then when something does go wrong at birth, then it must be congenital.	Bridging inference
6) Congenital heart problems are the ones that they are born with. Usually this doesn't happen, but sometimes, about 1 in 200 babies, are born with heart problems and I guess these problems are called congenital heart problems.	Paraphrase and bridging inference
7) When someone is born with a heart problem, it is called a congenital heart disease. This is fairly rare. I think that this is sometimes due to defective valves – and then the blood doesn't flow right.	Paraphrase, bridging inference, and elaboration

example constitutes a bridging inference because the statement clearly links the two sentences from the text. Hence, in our own research, we tend to focus on the relationship between the statement and the text, and in turn, the processes or strategies that contribute or lead to a self-explanation or think-aloud statement, rather than whether it is or is not an explanation per se.

Levels of Comprehension and Reading Processes

One dimension of our respective programs of research has been to assess the relationship between strategies revealed in verbal protocols (under instructions to self-explain or think aloud) and outcome measures of comprehension. This research is important to the primary thesis of this chapter because one would expect verbal protocols to be related to metacomprehension only if these strategies were differentially related to comprehension. In order to assess comprehension, we typically develop short-answer questions that require readers to access different aspects of their comprehension for the texts (e.g., memory for the explicit content or implied relationships between events). Many researchers, including us, assume that performance on comprehension questions after reading a text is reflective of the reader's level of understanding. One reason for our strong conviction regarding the relationship between reading strategies when producing verbal protocols and metacognition stems from the consistent finding that these strategies are correlated with performance on these measures, and in some instances in non-intuitive ways.

A basic assumption in reading comprehension research is that readers' understanding of the text can be described in terms of various interconnected levels (e.g., Kintsch, 1998). Most important here are the textbase and situation model levels. The textbase level of comprehension comprises memory for the gist of the text, in contrast to the situation model, which is formed from the inferences that the reader generates. These inferences may create links that reflect the relationships between separate ideas or sentences in the

text, or the inferences may comprise related knowledge that the reader conjures about the text. The readers' levels of comprehension are assumed to be reflected by readers' ability to answer certain types of questions. Text-based questions and recall primarily assess the reader's textbase level of comprehension. Text-based questions are those that require only memory or comprehension for single sentences or idea units within the text. For example, a typical question we have used in our own research to assess comprehension of the first sentence of the text targeted in Table 5.1, *A congenital disease is one with which a person is born*, is: *If a person has a congenital heart problem, at what stage of life did it most likely occur?* A correct response to this question indicates that the reader has understood the sentence. Proportion correct on this genre of questions is reflective of the likelihood that the reader has developed a coherent textbase representation of the entire text. In contrast, we might also ask, *How often does congenital heart disease occur?* This question would be called a *local bridging inference* question. Although subtle, correctly answering the question requires that the reader make the inference that *something going wrong in 1 in 200 cases* refers to the likelihood of congenital heart disease. Thus, readers' performance on this type of question would be more reflective of situation model level of comprehension.

If a reader performs relatively well on text-based questions and poorly on bridging inference questions, we assume that the reader has a fairly coherent textbase but a poor situation model. If a reader performs relatively well on both text-based and bridging inference questions, we assume that the reader has a coherent textbase and situation model. One question we have examined is whether the reader's level of comprehension is reflected in their verbal protocols. That is, is there a direct link between the cognitive processes engaged during reading and their level of comprehension?

One study that examined this issue was conducted by McNamara et al. (2006). The participants were 39 middle-school students entering the eighth and ninth grades (approximately 13 years old). In one part of the study, the students self-explained a text about heart disease and then answered comprehension questions. Non-parametric HLM analyses, called *Generalized Estimating Equation* (GEE) analyses, were conducted to examine the relationship between inferences (based on the text or prior knowledge) present in the self-explanations of the sections of text corresponding to the comprehension questions. The GEE analyses showed that there was a direct relationship between the strategies evident in the self-explanations and the students' performance on the questions. There was no relationship found between performance on the text-based questions and the presence of inferences in the self-explanations. Performance on these questions was driven primarily by strategies, such as paraphrasing. In contrast, performance on the bridging inference questions was directly linked to whether the student generated inferences when self-explaining the text corresponding to the questions. However, performance on these questions was also driven by students' prior knowledge of reading strategies. Thus, while the self-explanation protocols captured a good deal of the processes driving performance on the questions, students must have used additional reading and metacognitive strategies that were not evident in the self-explanations themselves.

Magliano and Millis have investigated this issue using think aloud rather than self-explanation (Magliano & Millis, 2003; Magliano, Trabasso, et al., 1999; Magliano, Millis, Ozuru, & McNamara, 2007; Magliano, Millis, et al., 2006). They had participants think aloud (either by verbalizing or typing thoughts) after targeted sentences requiring bridging inferences (as determined by a Causal Network Analysis; Trabasso, van den Broek, & Suh, 1989). They used computer-based approaches to assess reading strategies (Magliano et al., 2002; Millis, Magliano, Todaro, & McNamara, 2007). One approach involves using Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) to measure the semantic overlap between the reader's protocol and the current sentence (i.e., sentence that

was just read) to provide a measure of paraphrasing and the causally important prior text sentences to provide a measure of bridging inferences. In general, they have shown that the overlap scores between the protocols and prior discourse context are positively correlated with a number of comprehension measures (question answering performance, text recall), whereas the overlap scores for the current sentence are negatively correlated with these outcome measures. However, these relationships are contingent on the aspect of comprehension primarily assessed by the outcome measure (Magliano et al., 2007; Millis et al., 2006). For example, Magliano et al. (2007) found positive correlations between overlap scores for the causal antecedent sentences and both textbase and situation model question-answering performance. Hence, when a participant focused on causally related sentences, both levels of performance were enhanced. However, there was a negative correlation between overlap scores for the current sentence and textbase question performance. Thus, when readers focused on the current sentence, this was a signature of poor comprehension at the textbase level. As was the case with the textbase short-answer questions, Magliano et al. (2007) found a negative correlation between performance on the Nelson–Denny Reading Test and overlap with the current sentence. Thus, poor readers tended to focus more on the current sentence, rather than causally related sentences.

Presumably, the extent to which readers produce bridging explanations has implications on the quality of both textbase and situation model representations. These inferences enable readers to establish the implied causal relationships between text sentences and as such, it makes sense that there would be a positive and significant correlation between their production during thinking aloud and questions that tap these aspects of comprehension. It is reasonable to assume that the textbase would be strengthened when readers paraphrase text constituents. Thus, on the face of it, it may be counter-intuitive that there is a negative correlation between overlap scores with the current sentence and textbase questions. However, McNamara (2004) found that only correct paraphrases were positively correlated with comprehension question performance, and incorrect paraphrases negatively correlated with comprehension. Thus, the negative relationship between sentence overlap and textbase comprehension may be due to incorrect paraphrases. Also, a participant's tendency to paraphrase (over the use of other strategies) may be a signature of poor comprehension. Along these lines, Magliano and Millis (2003) argued that the high cosines calculated by LSA with the current sentence occur when readers paraphrase at the expense of other strategies. As such, excessive paraphrasing while thinking aloud may lead to a disjointed and incoherent textbase representation.

Best, Ozuru, and McNamara (2004) and Ozuru et al. (2004) found similar results with middle-school children. Analyses of self-explanations generated by these students indicated that skilled readers (measured by the Gates MacGinitie) produced better paraphrases (as opposed to simple repetition or topic identification) and were more likely to generate bridging inferences and knowledge-based elaborations. Better comprehenders were more likely to produce knowledge-building explanations that lent to a more global understanding of the text, linking various sentences from the text. Also, they were more likely to use a diversity of strategies. They did so regardless of their level of domain knowledge about the text. However, their higher quality inferences were often associated with “easy sentences.” This pattern implies that skilled readers are not necessarily good at solving problems associated with the comprehension of the difficult sentences. Rather they are good at actively relating different parts of texts to construct a global representation of the text.

In sum, the reading strategies that comprise self-explanation and think-aloud protocols are differentially correlated with performance on outcome measures of comprehension and reading skill. Magliano and Millis (2003; Gilliam et al., 2007) have taken the relationship a step further, using reading strategies identified in think-aloud protocols as a basis for

detecting individual differences in comprehension skill (Magliano & Millis, 2003; Millis et al., 2006). Magliano and Millis (2003) showed that skilled and less-skilled comprehenders, as measured by the Nelson–Denny Reading Test of comprehension, adopt different strategies when thinking aloud. Skilled readers tend to produce more bridging explanations and have higher overlap scores with the causal antecedent sentences than less-skilled comprehenders. Although less-skilled comprehenders do not produce more paraphrases than skilled readers, their overlap scores with the current sentences were greater than those of skilled readers.

Based on this research, Magliano and Millis have developed the Reading Strategy Assessment Tool (R-SAT; Gilliam et al., 2007; Magliano, Millis, The R-SAT Development Team, Levinstein, & Boonthum, 2007). R-SAT is an automated computer-based reading assessment tool designed to measure readers' comprehension and spontaneous use of reading strategies while reading narrative, history, and science texts. In the test, readers comprehend passages one sentence at a time and are asked either an indirect ("What are your thoughts regarding your understanding of the sentence in the context of the passage?") or a direct ("Wh-") question after pre-selected target sentences. The answers to the indirect questions are analyzed on the extent to which they contain words associated with reading strategies, such as paraphrasing, bridging, and elaborating. In addition, the answers to direct questions are coded on the extent to which they contain words in an ideal answer, which is a measure of passage comprehension. The R-SAT measure of strategy use and comprehension accounts for significantly more variance in performance on short-answer questions that tap textbase or situation model representation (unique variance explained = 27%) than the Gates–MacGinitie standardized test of reading comprehension (unique variance explained = 8%; Magliano et al., 2007).

Self-Explanation Training

Studying self-explanation and think aloud helps researchers to better understand the reading process. Importantly, self-explanation also helps readers to better understand what they are reading. Prompting learners to self-explain helps them to develop a deeper understanding of the material (e.g., Chi & Bassok, 1989; Chi et al., 1994; Chi & VanLehn, 1991; Magliano, Trabasso, et al., 1999; Trabasso & Magliano, 1996a, 1996b; VanLehn et al., 1992). Unfortunately, while this research revealed the benefits of self-explanation, it has also indicated that some readers do not naturally self-explain text and self-explain poorly when prompted to do so. These latter results prompted the question of whether readers can be trained to more effectively self-explain text, and in turn increase a student's metacognitive awareness of the appropriateness of various reading strategies that are used while self-explaining.

For that, McNamara and colleagues turned to the reading strategy literature. Specifically, can the quality of self-explanations and comprehension be improved by providing reading strategy instruction in combination with self-explanation, as opposed to self-explanation instruction alone? Indeed, a good deal of research has indicated that providing readers with instruction to use and practice reading strategies improves reading comprehension skills (Baker, 1996; Baumann, Seifert-Kessell, & Jones, 1992; Bereiter & Bird, 1985; Davey, 1983; Dewitz, Carr, & Patberg, 1987; Hansen & Pearson, 1983; Palincsar & Brown, 1984; Yuill & Oakhill, 1988; see also, McNamara, 2007, for a collection of chapters on this topic). Thus, our approach to this question has been to capitalize on the effectiveness of self-explanation as a learning strategy in combination with reading strategies (e.g., Magliano et al., 2005; McNamara, 2004; McNamara et al., 2004). The advantage of this approach is that it provides an effective intervention, and it also allows

us to examine change in strategy use over time. Thus, in this case, self-explanation is used as a vehicle for externalizing the reading process for the reader and helping the reader to engage in greater metacognitive reflection.

Two such interventions have been developed. The first is a live, or human-delivered, intervention called Self-Explanation Reading Training (SERT; McNamara, 2004; McNamara & Scott, 1999). The second is an automated version of SERT, called iSTART (Interactive Strategy Training for Active Reading and Thinking; McNamara et al., 2004; Levinstein et al., 2007). Studies have confirmed that there is a positive impact of SERT and iSTART on both reading strategies and comprehension for thousands of high school and college students (McNamara, 2004; McNamara et al., 2004, 2006; McNamara, O'Reilly, Rowe, Boonthum, & Levinstein, 2007; O'Reilly, Best, & McNamara, 2004; O'Reilly, Sinclair, & McNamara, 2004a, 2004b; O'Reilly, Taylor, & McNamara, 2006; Taylor, O'Reilly, Rowe, & McNamara, 2006; Taylor, O'Reilly, Sinclair, & McNamara, 2006).

Both SERT and iSTART provide students with instruction on how to more effectively self-explain text using five reading strategies: *monitoring comprehension* (i.e., recognizing comprehension failures and the need for remedial strategies), *paraphrasing* the text, making *bridging inferences* between the current sentence and prior text, making *predictions* about the subsequent text, and *elaborating* the text with links to what the reader already knows. For each strategy, a description of the strategy and examples of self-explanations using the strategies are provided. Comprehension monitoring is presented as a strategy that should be used all of the time and its emphasis in the interventions reflects the fact that we believe it to be critical for successful self-explanation. Paraphrasing is presented as a basis or jumpstart for self-explanation, but not as means for self-explaining the text. The remaining strategies are various forms of inferences (i.e., bridging, predictive, domain-specific knowledge, domain-general knowledge).

SERT training can be administered to small groups of students in about two hours. iSTART is delivered via the internet, using automated, interactive avatars to deliver one-on-one instruction and practice. Both SERT and iSTART begin with a brief *Introduction*, including definitions and examples of self-explanation and reading strategies. During the Introduction Module of iSTART, the trainee is engaged by a trio of *animated avatars*, including a *teacher* avatar (Julie) and two *student* avatars (Sheila and Mike), that interact with each other by motivating the intervention, providing examples, posing questions, and explaining the reading strategies (see Figure 5.1).

The *Introduction* is followed by the *Demonstration Module*. In SERT, students read a science text and watch a video of a student engaging in the process of self-explaining the text. The video is paused at certain points and the students are asked to identify the strategies used by the student in the video. In iSTART, students observe a student avatar (e.g., see Genie in Figure 5.2) who self-explains a text with feedback from a teacher avatar (e.g., Merlin). As in SERT, the students identify strategies in the avatar's explanation. In iSTART, the amount of scaffolding that the students receive depends on their reading skill level and their performance.

The final phase is *Practice*. In SERT, the students work with partners to practice the strategies, taking turns in reading aloud and explaining sections of text. In iSTART, students self-explain text, receiving feedback from an avatar (i.e., Merlin). The feedback is driven by computational algorithms that interpret the student's self-explanation (McNamara et al., 2007).

There were two purposes for the first study conducted on SERT (McNamara, 2004; McNamara & Scott, 1999). The first was to examine whether self-explanation and reading strategy training (i.e., SERT) would help readers to better explain text and learn more effective reading strategies, and by consequence, better understand challenging text. The

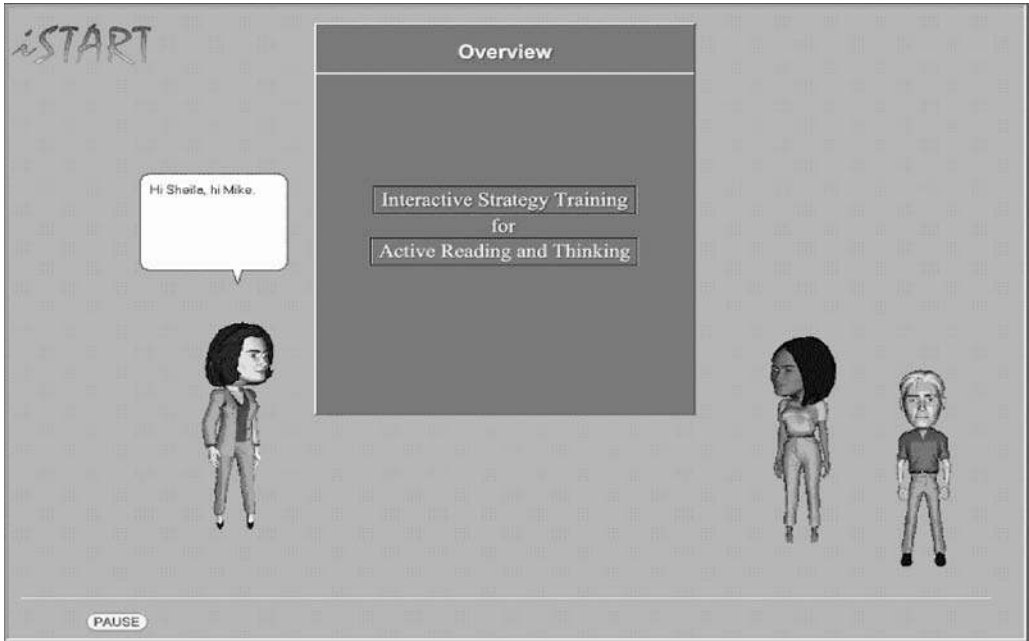


Figure 5.1 The iSTART Introduction Module comprises three animated agents including Julie the teacher (on the left) and two students, Sheila and Mike.

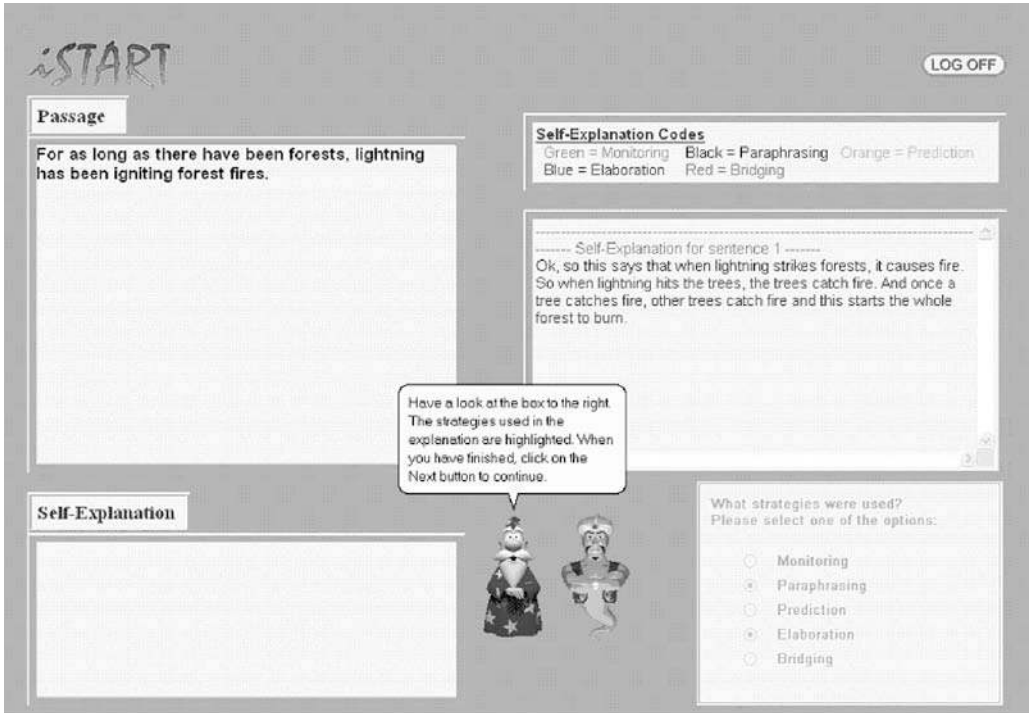


Figure 5.2 During the iSTART Demonstration Module, two agents (Merlin and Genie) demonstrate the use of self-explanation using a science text and the trainee identifies the strategies being used by the agents.

second purpose was to examine whether the effectiveness of SERT depended on some facet of readers' abilities. Assuming SERT was effective, then the secondary goal was to identify an individual difference measure that could indicate whether this type of training would be more or less effective for particular individuals.

McNamara (2004; McNamara & Scott, 1999) first examined the effects of SERT with 42 adult readers, half of whom received SERT training. Each participant in the SERT condition received individual training, practiced using the strategies with four texts, and watched four videos of students self-explaining those texts. After training, all participants self-explained a low-cohesion cell mitosis text (used in McNamara, 2001). Three individual difference measures were examined. First, working memory capacity was assessed using the reading span test developed by Engle and his colleagues (e.g., Engle, Cantor, & Carullo, 1992; La Pointe & Engle, 1990). Second, reading skill was assessed using the Nelson-Denny Reading Test. Third, participants answered open-ended questions that assessed their prior knowledge of cells and cell mitosis (i.e., the topic of the target text).

The results showed that neither reading skill nor working memory capacity were correlated with performance either during training (i.e., for the four training texts) or after training (i.e., for the target cell mitosis texts), nor did they interact with the effects of training condition. Only participants' prior knowledge of cells was predictive of comprehension. Specifically, the results indicated that SERT was not effective for readers with more prior knowledge, whereas it was highly effective for low-knowledge readers. Few benefits were expected for high-knowledge readers because these readers automatically use their knowledge to bridge conceptual gaps in the text. In contrast, reading strategy training helped the low-knowledge reader to use logic and common sense rather than domain-relevant prior knowledge to fill in conceptual gaps. While prior knowledge may be the most direct and natural way to resolve cohesion gaps, the reader may be able to "work harder" to understand the text by generating more logic-based and text-based inferences. If that is the case, however, benefits of strategy training should depend on the level of understanding assessed by the comprehension questions.

Indeed, McNamara (2004) found an interaction between the benefit of SERT and the type of comprehension question. As Figure 5.3 illustrates, the benefits only appeared on the more text-based measures of comprehension, as opposed to the more knowledge-demanding situation model comprehension questions. This is because the development of

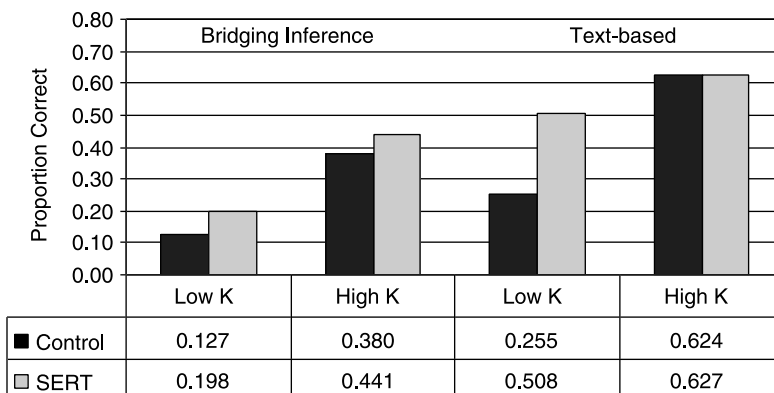


Figure 5.3 Comprehension of the post-training text about cell mitosis, self-explained by both control and SERT participants in McNamara (2004) as a function of condition, knowledge, and question type. This three-way interaction shows reliable effects of training only for low-knowledge participants' performance on text-based questions (Cohen's $d = 1.92$).

a coherent situation model for a text is highly dependent on having sufficient prior knowledge. According to bridging inference questions, only prior domain knowledge helped the readers to make the necessary inferences. In contrast, text-based questions revealed an effect of training for the low-knowledge readers. SERT was most effective for students who had less prior knowledge about the text domain. This training provided students with strategies they could use while reading, which effectively compensated for their lack of domain knowledge. Further supporting this interpretation, protocol analyses indicated that low-knowledge readers who had been trained with SERT relied on their common sense and logic to understand the text.

While this initial study provided good evidence for the effectiveness of SERT, it did not provide an indication as to who may benefit most from SERT. While the prior knowledge measure was successful, prior knowledge tends to vary according to topic and interests of the reader. There are some readers who are relatively low knowledge in all topics, and there are some readers who tend to know more about many topics, but there are few readers that are high knowledge in all topics. Thus, while the interaction between prior knowledge and training furthered our theoretical understanding of reading comprehension, it was not clear that there were clear practical implications from the result. In sum, the results indicated that self-explanation and reading strategy training may be effective for those who by chance were assigned to read a challenging text about a topic with which they were only mildly familiar. Hence, one goal of the subsequent research was to discover if another ability measure, other than prior knowledge, successfully identified those who would or would not benefit from SERT. One of the variables we have examined is metacognition.

There is some limited evidence that strategy training could have an impact on metacognition. Magliano, Little, and Graesser (1993) explored whether brief exposure to comprehension and reading strategies impacts the calibration of comprehension. They briefly exposed readers to strategies to promote discourse comprehension (e.g., questioning, summarizing, predicting, strategic inference generation) or strategies to promote word or sentence processes (e.g., orthographic processing, phonemic processing, syntactic processing). After exposure to the strategies, participants were told to use the strategies while reading four texts. After reading each text, they made judgments regarding their expected performance on a comprehension test and then answered questions about the text. Although exposure to the deep comprehension strategies did not influence comprehension calibration accuracy, exposure to the shallow strategies actually led to a temporary decrease in accuracy. These results must be qualified by the limitation that participants were only briefly exposed to the strategies. However, the results clearly suggested that it is worth while to further explore the extent to which strategy training improves or hinders readers' ability to gauge their level of understanding of a text.

Metacognition, Comprehension, and Self-Explanation

It seems intuitive that a measure of metacognition should yield a good indicator on who needs self-explanation and reading strategy training, and who does not. That is, if better readers tend to exhibit greater metacognitive skills (see McKeown & Beck; Harris, Graham, Brindle, & Sandmel; and Paris in this volume), and if self-explanation and reading strategies rely to some extent on metacognition, then a measure of metacognition should yield a good measure of who needs self-explanation training. However, there have been several challenges. The most significant challenge has been how to measure metacognition.

For the most part, we have attempted to measure only one facet of metacognition: the knowledge and use of comprehension strategies. One measure that we have explored is the

Metacomprehension Strategy Index (MSI; Schmitt, 1990). The MSI is a 25-item multiple-choice questionnaire that assesses students' knowledge of reading strategies used before, during, and after reading a text, such as predicting and verifying, previewing, purpose setting, self-questioning, drawing from background knowledge, and summarizing. This questionnaire assesses students' awareness of strategies as they pertain to reading comprehension, as opposed to a general or all-encompassing form of metacognition. Some examples are provided in Table 5.2.

Table 5.2 Examples of Questions in the Metacomprehension Strategy Index

Before I begin reading, it is a good idea to:

- a. guess what the chapter is about on the basis of the title and pictures.
- b. see how many pages are in the chapter.
- c. estimate how long it will take to read the chapter.
- d. read the middle part of the chapter first.

While I read, it is a good idea to:

- a. check to make sure that I understand what I am reading.
- b. mentally repeat the sentence to myself.
- c. underline the sentence.
- d. try to focus on only the current sentence.

After I read, it is a good idea to:

- a. try to figure out which sections I didn't understand.
 - b. count how many words I didn't understand in the chapter.
 - c. try to remember the table of contents.
 - d. memorize the chapter review questions.
-

We also developed a new measure, called the Reading Strategy Checklist, which included 32 statements. The reader was asked to place a check by each strategy that was typically used by the reader. The strategies included ineffective reading strategies, superficial strategies, moderately effective strategies, repair strategies, inference strategies, comprehension monitoring, and previewing. Some examples are provided in Table 5.3.

Table 5.3 Examples of Statements in the Reading Strategy Checklist

- I skip sections that are too difficult for me.
 - I have a hard time concentrating while I read.
 - I often don't read the textbook because I don't have time.
 - I think about how the pictures relate to the information in the chapter.
 - I check to see how many pages are in the chapter that I am going to read.
 - I read the chapter one time, and I don't usually need to read it again.
 - I think about how ideas relate to each other in the chapter.
 - I stop to review the main points to see if I understand what I've read so far.
-

The first study that examined these two measures included 1651 high-school students from four high schools, distributed across three states (Virginia, Kentucky, and Georgia).^{*} This study confirmed that the internal consistency of the Reading Strategy Checklist was fairly good. For example, there was a -0.62 correlation between good reading strategies and ineffective reading strategies. However, the correlations between the MSI and the checklist total scores and subtest scores were low (between -0.16 and 0.21). The correlations were too low to indicate that the two measures were assessing the same construct.

^{*} The results of this study were reported in O'Reilly and McNamara (2002, 2007). However, the majority of the results that we report here were not included in the final versions of those papers.

Table 5.4 Correlations of Reading Skill (as Measured by the Gates), Prior Science Knowledge, MSI, and Reading Strategy

<i>Individual measure</i>	<i>Course grade</i>	<i>VA Standards of Learning test</i>	<i>Open-ended comprehension</i>	<i>Multiple-choice comprehension</i>
Reading skill	0.24	0.58	0.64	0.53
Science knowledge	0.25	0.59	0.55	0.51
MSI	0.20	0.15	0.26	0.24
Reading strategy checklist*	-0.13	-0.07	-0.05	-0.03

* The reading strategy checklist consisted of measures of student performance, including course grades, performance on the Virginia Standards of Learning test for biology, and an experimenter-created measure of text comprehension that included open-ended and multiple-choice questions

The correlation between the checklist measure and performance on the Gates Reading Comprehension Test and a measure of prior science knowledge were also very low (i.e., -0.06 and -0.03 , respectively). In contrast, the MSI showed correlations of 0.30 with reading skill and 0.28 with prior science knowledge.

Table 5.4 further shows the correlations of reading skill (as measured by the Gates), prior knowledge of science, MSI, and the checklist with measures of student performance, including course grades, performance on the Virginia Standards of Learning test for biology, and an experimenter-created measure of text comprehension that included open-ended and multiple-choice questions. The results indicated that the Reading Strategy Checklist did not correlate with any measures of student performance. The MSI showed relatively low correlations, particularly as compared with the measures of reading skill and prior knowledge. Thus, the results for both the MSI and the checklist, in terms of their ability to predict student performance in science, were largely disappointing. In addition, while reliability was acceptable for the Reading Strategy Checklist, its construct validity was very low. As a consequence, we abandoned the checklist, and did not use it in any further studies.

Subsequent results in further studies that included the MSI have been largely disappointing. For example, another study with 1025 high-school students from two schools in Virginia and one school in Kentucky, yielded similar results to the previous one. The MSI correlated 0.37 and 0.32 with reading skill and prior knowledge and correlated 0.20 with reading comprehension tests. In most studies where the MSI was somewhat predictive of the effectiveness of training, the Gates reading comprehension test was usually a better predictor and usually showed higher correlations with comprehension performance measures.

In his thesis project, Perry (2008) provided additional evidence that the MSI is poorly correlated with comprehension and strategy use. He assessed the extent to which aspects of the reader and reading strategies were correlated with comprehension and the extent to which readers' attributes were correlated with the use of strategies. He measured working memory capacity, metacognitive awareness (with the MSI), need for cognition, vocabulary knowledge, and exposure to texts, in a sample of college students. Participants read a series of texts on the computer. Perry assessed the extent to which participants paraphrased, bridged, and elaborated while typing verbal protocols under think aloud instructions. Comprehension for these texts was assessed via online short-answer comprehension questions. These questions were primarily why-questions regarding the sentence that was just read. Perry conducted a series of regression analyses to assess the relationship between the attributes of the reader and reading strategies with comprehension. With respect to readers' attributes, only vocabulary knowledge and working memory span were significant

positive predictors of comprehension performance, whereas the extent to which readers paraphrased, bridged, and elaborated were all significant and positive predictors of comprehension. Moreover, reading strategies accounted for 22% of the unique variance in comprehension performance, whereas readers' attributes accounted for only 9%. Apparently, the successful use of reading strategies is more predictive of comprehension than the attributes that the reader brings into the reading situation. Moreover, these findings are consistent with the research discussed earlier showing that the MSI is not strongly correlated with comprehension performance. Indeed, Perry (2008) found that MSI was not significantly correlated with any of the strategies. Although the Need for Cognition scale was positively and significantly correlated ($r = 0.18$) with the extent that the participants produced bridging inferences. The only measure that was consistently correlated of strategy use was vocabulary knowledge ($r = 0.47$ for paraphrasing, $r = 0.24$ for bridging, and $r = 0.27$ for elaboration).

There has been only one exception to the trend that the MSI is poorly correlated with strategy use, comprehension, and response to self-explanation training. The study conducted by McNamara et al. (2006) with middle-school students discussed earlier also investigated the effectiveness of iSTART (i.e., the automated version of SERT). The results showed that the effectiveness of training on their ability to comprehend science text depended on both their strategy knowledge (as indicated by the MSI), as well as the type of comprehension question. Among students who performed less well on the MSI (low strategy knowledge), those given iSTART training performed substantially better than the control participants, but only on the text-based questions (Cohen's $d = 1.00$). In contrast, among the high-strategy knowledge students, those given iSTART training performed better on the bridging-inference questions (Cohen's $d = 1.04$). As such, this study indicated that the MSI was not indicative of *who* would benefit from training, but rather *how* they would benefit from training: The low-strategy knowledge students improved in terms of text-based comprehension (for that text), whereas the high-strategy knowledge students improved at the level of the situation model. These results have been replicated by Magliano et al. (2005), but using a measure of reading skill (the Nelson–Denny Reading Test), rather than with a measure of strategy knowledge, or metacognition.

By and large, the results of the majority of the studies we have conducted with the MSI have indicated that it correlates moderately with comprehension and may be predictive of training gains with younger students. However, for the most part, the results have been largely disappointing. Likewise, we have had disappointing results using the the Metacognitive Awareness of Reading Strategies Inventory (MARSİ; Mokhtari & Reichard, 2002).

Despite our limited success using standardized measures of metacognitive awareness to assess the relationship between reading strategies and comprehension, we believe that metacognitive awareness is critical for the successful use of comprehension strategies. One basis of this belief stems from the dynamic nature of self-explanation and thinking aloud. As we will describe, readers alter their use of strategies based on the demands of the texts, their goals, and their comprehension failures and success. Although we do not have strong evidence yet, we believe that a sense of metacognitive awareness enables readers to modulate their strategy use. Our goal in the following section is to present evidence that this relationship warrants further investigation.

The Dynamic Nature of Self-Explanation and Reading Strategies

Self-explanation, by its very nature, calls on the reader to assess comprehension, as it relates to a host of factors, including what the reader knows about the information in the text, the difficulty of the text, the importance of the text, the reader's comprehension goals,

and the reader's awareness of the strategies that can be used to meet those goals. Thus, self-explanation involves dynamically changing processes by which readers engage in whatever strategies are possible for that reader and deemed important for constructing meaning at any given point in a text. For example, explanations of narrative text are most appropriate when the text describes a causal process and the reader is asked to understand an event that is part of that process. However, not all texts or sentences describe events that exist on a causal chain. For example, when readers generate a self-explanation for a topic sentence, it may be most appropriate to use the content of the sentence to anticipate future text content. Additionally, generating explanations may not be feasible or appropriate given a reader's level of comprehension. If, for example, a sentence contains unfamiliar words, the most appropriate strategy may be to rephrase it using more familiar words (i.e., paraphrasing) as opposed to generating knowledge-based inferences.

Hence, in our research, we have typically described self-explanations in terms of their relationship and appropriateness to the target text, individual sentences within the text, and the reader's level of comprehension, rather than their relationship to higher-level processes (e.g., explanation, reasoning, problem solving). In turn, the relation between the text and the protocol typically corresponds to metacognitive processes or reading strategies that are likely being engaged by the reader. The nature of this process will be heavily influenced by the extent that the reader is aware of and applies strategies that are appropriate to the texts, individual sentences, and the reader's comprehension goals.

Our ultimate goal is to better understand comprehension processes by examining the traces of these processes in protocols that are generated in the context of tasks such as self-explanation and think aloud. For example, "I don't understand this" expresses the reader's lack of understanding, but also indicates that the reader is engaging in comprehension monitoring. The third example in Table 5.1 focuses on each of the two sentences in the target text separately and rephrases each of the sentences into the reader's own words. From this, we can characterize the protocol as consisting of paraphrases; and in turn, we can infer that the reader is focusing on the text and is not generating inferences concerning the relations between concepts expressed in the text or relations between the text and what the reader already knows. Hence, the reader is primarily engaging in text-based processing, in contrast to building a deeper understanding of the text (Coté & Goldman, 1999). This contrasts with Examples 5 and 6, where the reader expresses the relationship between the two sentences, which requires making inferences that bridge or link the sentences.

Analyses of self-explanations and think-aloud protocols provide some indication of what strategies readers tend to use while reading, and which strategies are used in combination, and which are not. For example, McNamara (2004) examined the frequency of strategies used by adult readers who did and did not receive prior training to self-explain while self-explaining science text. McNamara (2004) found that the most predominant strategy used by participants was paraphrasing. Accurate paraphrases were evident in 76% of the explanations and incorrect paraphrases were present in 13% of the explanations. The next most likely strategy was making bridging inferences linking the target sentence to a prior sentence: Correct bridging inferences were found in 38% and incorrect bridging inferences were present in 11% of the explanations. The least likely strategy to occur was prediction. This result replicated findings reported by Magliano et al. (1993) who found evidence using a lexical decision task that readers generate few predictions during the normal course of reading. Likewise, Magliano, Trabasso, et al. (1999) found that when readers used think aloud with narrative text, predictions were unlikely unless the reader was instructed to make predictions.

Magliano and his colleagues have found that college students produce knowledge-based explanations more frequently than any other strategy when reading simple narrative texts (Magliano, Trabasso, et al., 1999; Trabasso & Magliano, 1996a, 1996b). These results contrast with those reported by McNamara (2004), presumably because some kinds of elaboration may be more prevalent for narrative texts than for science texts. Differences as a function of genres such as narrative and science likely occur because of differences in the amount of relevant background knowledge possessed by the participants. In our studies with college students reading science texts, they typically do not possess extensive background knowledge associated with the topics described in the text, whereas these students have greater knowledge of the types of social situations that arise in the stereotypical narratives that are used in our research. Thus, strategy use is likely to depend on the reader's abilities as well as the type of text.

We have also found systematic relationships between the co-occurrence of different strategies (McNamara, 2004; Todaro, Magliano, Millis, Kurby, & McNamara, 2009). McNamara (2004) found that paraphrases and bridging inferences were likely to co-occur. Indeed, 84% of the self-explanations containing bridging inferences also contained paraphrases. In contrast, strategies that went beyond the text-based information, such as elaboration and prediction, were less likely to co-occur with paraphrasing. Similarly, Perry (2008) found moderate but significant correlations between the extent that college students generated paraphrases and bridging inferences ($r = 0.32$) when producing verbal protocols under think-aloud instructions. However, paraphrasing was not correlated with elaboration. Both Perry and McNamara also found that there was no significant correlation between readers' generation of bridging inferences and elaborations.

Paraphrasing may be a highly prevalent strategy and co-occur with bridging inferences because it serves as an "anchor" for these inferences when producing verbal protocols. That is, when reading to explain how the current sentence is related to the prior discourse, readers tend to paraphrase the current sentence to form a foundation. As will be discussed below, the lack of a strong relationship between bridging and elaboration may be because the features of the text that afford these inferences tend not to co-occur (Magliano, Zwaan, & Graesser, 1999; Magliano, Trabasso, et al., 1999).

Magliano, Trabasso, et al. (1999) provided evidence for a dynamic perspective of reading strategies in the context of a think-aloud task. They asked college students to think aloud after every sentence while reading simple narrative texts. Discourse analyses of the texts were conducted to identify factors that were likely to influence the probability that readers generated explanations based on the prior texts or world knowledge. They identified two factors that were related to the extent to which readers generated explanations based on prior text content or world knowledge. One important factor was the presence or absence of causal antecedents in the text. The importance of causal antecedents in texts follows from research showing that relationships inferred by a reader are constrained by what is afforded by the text (Trabasso et al., 1989; van den Broek, 1990). Magliano and colleagues identified causal relationships between story elements in the text on the basis of causal network analysis (Trabasso et al., 1989). This analysis provided an estimate of the number of causal antecedents and causal consequences in each text. They also identified whether a sentence introduced new entities (i.e., new characters or objects), and hypothesized that the presence of these new entities would activate new and potentially relevant world knowledge that could be used while thinking aloud.

Magliano, Trabasso, et al. (1999) found that the presence of causal antecedents in prior text was positively correlated with the number of explanations based on prior text content. In contrast, the lack of causal antecedents was negatively correlated with more knowledge-based (elaborative) inferences. Conversely, Magliano et al. found a positive correlation

between the introduction of new entities and the production of knowledge-based explanations, whereas this variable was negatively correlated with the production of text-based explanations. Thus, for relatively familiar and easy text, skilled (adult) readers are more likely to make inferences that link back to prior text when the text affords those inferences. When it does not, and when the text introduces new elements, the reader is more likely to reach outside the text, making knowledge-based inferences.

Todaro et al. (2009) conducted a similar study, but in the context of science texts. They assessed factors that influenced the availability of information from the prior text and world knowledge that could be used for inference and strategic processes. They had participants type their thoughts (via think-aloud instructions) after each sentence. They classified the content words (i.e., nouns, pronouns, verbs, adverbs, and adjectives) in terms of whether they came from the current sentence, prior text context, or world knowledge. Like Magliano, Trabasso, et al. (1999), they conducted discourse analyses of the text (i.e., a causal network and argument overlap analyses) to assess the presence of text-based factors that should influence the availability of information from the prior discourse and world knowledge. However, in addition to these analyses, they computed a measure of the extent to which a reader's relevant background knowledge with the topic of a text (e.g., the development of thunderstorms, heart disease) overlapped with the sentences contained in the text. Specifically, they had participants write down as much as they knew about the topics discussed in their text. They then used Latent Semantic Analysis (Landauer & Dumais, 1997; Landauer, McNamara, Dennis, & Kintsch, 2007) to compute a knowledge-resonance score, which was the cosine between each sentence of a text and the prior knowledge protocol associated with that text. The knowledge-resonance score measured the extent to which a reader's world knowledge overlapped with each sentence of the text.

Todaro et al. (2009) conducted a series of multiple regression analyses to determine factors that influence the use of content words from the current sentence, prior text, and world knowledge. Consistent with Magliano, Trabasso, et al. (1999), they found that factors that were theoretically associated with establishing relations between sentences (e.g., number of causal antecedents) were positively correlated with the inclusion of words from the prior discourse context, but negatively correlated with the use of words from world knowledge. Conversely, factors that were theoretically associated with elaboration (introduction of new argument nouns and knowledge overlap scores) were positively correlated with the inclusion of world knowledge words, but negatively correlated with the inclusion of prior text words.

These studies paint a consistent picture that is relevant to our perspective of self-explanation as a metacognitive process. Specifically, text-based and reader variables are differentially associated with the strategies that we conceptualize as composing self-explanation. The presence of these factors varies across sentences in a text, and therefore, the appropriate strategies for self-explanation vary as well. Readers will be more or less effective when self-explaining to the extent that they modulate the strategies they use across sentences.

Conclusions

We began this chapter with the claim that metacognitive processes are critical for successful self-explanation and for strategic reading in general. However, we have described several studies that have failed to establish that metacognition has a strong relationship with strategy use or even with various levels of comprehension. These studies adopted an approach of using standardized measures to assess individual differences in various dimensions of metacognition, such as the reader's awareness of and knowledge of reading

strategies. It may be the case that these assessment tools do not tap the aspects of metacognition that are most associated with the successful use of reading strategies. Presumably, if we had the right measure, then we would be better able to assess this relationship. However, the fact that we show relatively modest correlations with well-accepted measures of various levels of comprehension suggests that the measures may be flawed or inappropriate for college and high-school students. We advocate a slightly different perspective on this issue. Given the dynamic nature of comprehension and strategy use, relatively static standardized measures of metacognition will not adequately assess the extent to which readers are successful at modulating their strategy use given their goals, comprehension level, and the demands of the text.

The research presented here suggests that verbal protocols may be an alternative to standardized measures for exploring the role of metacognition in comprehension and strategy use. Although other researchers have explored the role of metacognition in comprehension via verbal protocols (Pressley & Aflerbach, 1995), none have explored the dynamic nature of strategy use in a manner consistent with our research (Magliano, Trabasso, et al., 1999; Magliano, Zwaan, et al., 1999; McNamara, 2004; Todaro et al., 2009). Specifically, we have used principled, theory-driven discourse analysis to identify the presence of factors that should lead skilled readers to engage in specific inferential and strategic processes. This research suggests that self-explanation is a multidimensional process that consists of many different strategies. Importantly, the appropriateness of strategies and combinations of strategies varies as a function of the readers' abilities and goals, the text, and the particular sentence. Not all strategies will be appropriate or effective in constructing a coherent understanding at any given point in the text.

Although Perry (2008) failed to show a strong relationship between dimensions of metacognition and the use of reading strategies when thinking aloud, we believe that this relationship warrants further investigation. Given that we advocate a constructionist perspective of discourse comprehension (Graesser et al., 1994), we believe that inference and reading strategies are under the control of the reader. As such, the reader must be able to adequately assess their level of comprehension, the demands of the texts, and the appropriateness of reading strategy at their disposal.

The specific nature of the relationship between metacognition and strategy use is unclear at this time. It may very well be the case that a variety of metacognitive processes are necessary for readers to successfully read strategically. It is also equally possible that metacognitive awareness emerges when readers engage in processes, such as self-explanation. That is an important assumption of self-explanation training (McNamara, 2004). Nonetheless, this relationship warrants further investigation.

In this chapter, we have described our efforts to develop intelligent systems that assess comprehension and promote self-explanation and reading strategies. These endeavors would benefit from a deeper understanding of the relationship between metacognition and reading strategies. Our research on gains from self-explanation have indicated that considering both the difficulty of the text and the abilities of the reader are critical in maximizing gains from self-explanation as well as training in reading strategies. Thus, one aspect of our current research has been directed toward finding that magic zone of proximal development. The overarching notion is to develop techniques whereby we can match a particular reader to a particular text such that the text is just difficult enough to push the reader beyond the zone of comfort, but not too difficult that nothing can be understood. This is a tricky endeavor in that we believe that training can be maximized if we can match texts to the needs and skills of a reader so that we can maximize their assessment of their comprehension and the appropriateness of strategies that they have mastered to enable them to achieve a coherent sense of comprehension.

With respect to assessment tools, R-SAT currently provides a measure of overall strategy use (Gilliam et al., 2007), but does not provide information regarding the extent to which readers use strategies that are appropriate given the features of the text. As such, R-SAT does not capitalize on the extent to which the verbal protocols reflect the dynamic use of strategies. One could construct a text where discourse analyses tools are used to determine which strategies are most appropriate, given the nature of the text. One could then assess whether students change their strategies in accordance to these demands. Presumably, doing so would make R-SAT more sensitive to individual differences in the metacognitive awareness of reading strategies and in turn, make it a better test of reading strategies.

The fundamental claim of this chapter is that reading is a dynamic process that fluctuates as an interaction between various factors. The dynamic nature of reading is an issue that warrants further investigation. Indeed, one of our current research goals is to further investigate the extent to which there are individual differences in how students dynamically shift between comprehension strategies. We expect that the strategies readers use at particular places in the reading process vary as a function of the text and sentence characteristics, as well as the readers' awareness of the reading process, knowledge of strategies, reading skill, and importantly, prior domain knowledge. Pinpointing the nature of these interactions will further our understanding of metacognition, reading comprehension, and the cognitive processes necessary to successfully learn from text.

References

- Baker, L. (1996). Social influences on metacognitive development in reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 331–352). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Baumann, J. F., Seifert-Kessell, N., & Jones, L. A. (1992). Effect of think aloud instruction on elementary students' comprehension monitoring abilities. *Journal of Reading Behavior*, *24*, 143–172.
- Bereiter, C. & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction*, *2*, 131–156.
- Best, R., Ozuru, Y., & McNamara, D. S. (2004). Self-explaining science texts: Strategies, knowledge, and reading skill. In Y. B. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of the sixth International Conference of the Learning Sciences: Embracing diversity in the Learning Sciences* (pp. 89–96). Mahwah, NJ: Erlbaum.
- Chi, M. T. H. & Bassok, M. (1989). Learning from examples via self-explanations. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 251–282). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Chi, M., Bassok, M., Lewis, M. W., Reimann, R., & Glaser, R. (1989). Self-explanation: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182.
- Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVanher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477.
- Chi, M. T. H. & VanLehn, K. (1991). The content of physics self-explanations. *Journal of the Learning Sciences*, *1*, 69–106.
- Coté, N. & Goldman, S. R. (1999). Building representations of informational text: Evidence from children's think-aloud protocols. In H. van Oostendorp & S. R. Goldman, *The construction of mental representations during reading* (pp. 169–193). Mahwah, NJ: Erlbaum.
- Davey, B. (1983). Think aloud: Modeling the cognitive processes of reading comprehension. *Journal of Reading*, *27*, 44–47.
- Dewitz, P., Carr, E. M., & Patberg, J. P. (1987). Effects of inference training on comprehension and comprehension monitoring. *Reading Research Quarterly*, *22*, 99–121.
- Engle, R., Cantor, J., & Carullo, J. (1992). Individual differences in working memory and comprehension: A test of our hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 972–992.

- Gilliam, S., Magliano, J. P., Millis, K. K., Levinstein, I., & Boonthum, C. (2007). Assessing the format of the presentation of text in developing a Reading Strategy Assessment Tool (R-SAT). *Behavior Research Methods, Instruments, & Computers*, 39, 199–204.
- Graesser, A. & Clark, L. (1985). *Structures and procedures of implicit knowledge*. Norwood, NJ: Ablex Pub. Corp.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371–395.
- Hansen, J. & Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology*, 75, 821–829.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Press.
- Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.
- Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). (Eds.). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.
- LaPointe, L. B. & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 1118–1133.
- Levinstein, I. B., Boonthum, C., Pillarisetti, S. P., Bell, C., & McNamara, D. S. (2007). iSTART 2: Improvements for efficiency and effectiveness. *Behavior Research Methods*, 39, 224–232.
- Magliano, J. P., Little, L. D., & Graesser, A. C. (1993). The impact of comprehension instruction on the calibration of comprehension. *Reading Research and Instruction*, 32, 49–63.
- Magliano, J. P. & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure. *Cognition and Instruction*, 21, 251–283.
- Magliano, J. P., Millis, K. K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies*, (pp. 107–136). Mahwah, NJ: Erlbaum.
- Magliano, J. P., Millis, K. K., The R-SAT Development Team, Levinstein, I., & Boonthum, C. (2007). Assessing comprehension strategies and performance with the Reading Strategies Assessment Tool (R-SAT). An unpublished manuscript.
- Magliano, J. P., Todaro, S. Millis, K. K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S. (2005). Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research*, 32, 185–208.
- Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processes during comprehension. *Journal of Educational Psychology*, 91, 615–629.
- Magliano, J. P., Wiemer-Hastings, K., Millis, K. K., Muñoz, B. D., & McNamara, D. S. (2002). Using latent semantic analysis to assess reader strategies. *Behavior Research Methods, Instruments, & Computers*, 34, 181–188.
- Magliano, J. P., Zwaan, R. A., & Graesser, A. C. (1999). The role of situational continuity in narrative understanding. In S. R. Goldman & H. van Oostendorp (Eds.), *The construction of mental representation during reading* (pp. 219–245). Mahwah, NJ: Erlbaum.
- McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55, 51–62.
- McNamara, D. S. (2004). SERT: Self-Explanation Reading Training. *Discourses Processes*, 38, 1–30.
- McNamara, D. S. (2007). IIS: A marriage of computational linguistics, psychology, and educational technologies. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the twentieth International Florida Artificial Intelligence Research Society Conference* (pp. 15–20). Menlo Park, California: The AAAI Press.
- McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2004, May). *Using LSA and word-based measures to assess self-explanations in iSTART*. Invited paper presented at the Latent Semantic Analysis Workshop, University of Colorado, Boulder, CO.
- McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007). Evaluating self-explanations in iSTART: comparing word-based and LSA algorithms. In T. Landauer, D. S. McNamara,

- S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 227–241). Mahwah, NJ: Erlbaum.
- McNamara, D. S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, *34*, 147–171.
- McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I. B. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 397–421) Mahwah, NJ: Erlbaum.
- McNamara, D. S. & Scott, J. L. (1999). Training reading strategies. *Proceedings of the twenty-first Annual Meeting of the Cognitive Science Society* (pp. 387–392). Hillsdale, NJ: Erlbaum.
- Millis, K., Magliano, J. P., & Todaro, S. (2006). Measuring discourse-level processes with verbal protocols and latent semantic analysis. *Scientific Studies of Reading*, *10*, 251–283.
- Millis, K. K., Magliano, J. P., Todaro, S., & McNamara, D. S. (2007). Assessing and improving comprehension with Latent Semantic Analysis. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *LSA: A road to meaning* (pp. 207–225). Mahwah, NJ: Erlbaum.
- Mokhtari, K. & Reichard, C. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, *94* (2), 249–259.
- O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-explanation reading training: Effects for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1053–1058). Mahwah, NJ: Erlbaum.
- O'Reilly, T. & McNamara, D. S. (2002, June). *What's a science student to do?* Poster presented at the 12th annual meeting of the Society for Text and Discourse, Chicago, IL.
- O'Reilly, T. & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional “high-stakes” measures of high school students' science achievement. *American Educational Research Journal*, *44*, 161–196.
- O'Reilly, T., Sinclair, G. P., & McNamara, D. S. (2004a). iSTART: A web-based reading strategy intervention that improves students' science comprehension. In Kinshuk, D. G. Sampson, & P. Isaías (Eds.), *Proceedings of the IADIS International Conference Cognition and Exploratory Learning in Digital Age: CELDA 2004* (pp. 173–180). Lisbon, Portugal: IADIS Press.
- O'Reilly, T. P., Sinclair, G. P., & McNamara, D. S. (2004b). Reading strategy training: Automated versus live. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1059–1064). Mahwah, NJ: Erlbaum.
- O'Reilly, T., Taylor, R. S., & McNamara, D. S. (2006). Classroom based reading strategy training: Self-explanation vs. reading control. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1887). Mahwah, NJ: Erlbaum.
- Ozuru, Y., Best, R., & McNamara, D. S. (2004). Contribution of reading skill to learning from expository texts. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1071–1076). Mahwah, NJ: Erlbaum. *Computing Research*, *34*, 147–171.
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction*, *2*, 117–175.
- Perry, P. J. (2008). *The role of general reading resources and discourse strategies in reading comprehension*. Unpublished Masters Thesis, Northern Illinois University.
- Pressley, M. & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schmitt, M. (1990). A questionnaire to measure children's awareness of strategic reading processes. *Reading Teacher*, *43*, 454–461.
- Taylor, R. S., O'Reilly, T., Rowe, M., & McNamara, D. S. (2006). Improving understanding of science texts: iSTART strategy training vs. web design control task. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2234). Mahwah, NJ: Erlbaum.
- Taylor, R., O'Reilly, T., Sinclair, G., & McNamara, D. S. (2006). Enhancing learning of expository science texts in a remedial reading classroom via iSTART. *Proceedings of the 7th International Conference of Learning Sciences* (pp. 765–770). Bloomington, Indiana.

- Todaro, S. A., Magliano, J. P., Millis, K. K., Kurby, C. A., & McNamara, D. S. (2009). Availability of information sources during reading. *Discourse Processes*.
- Trabasso, T. & Magliano, J. P. (1996a). Conscious understanding during text comprehension. *Discourse Processes*, 21, 255–288.
- Trabasso, T. & Magliano, J. P. (1996b). How do children understand what they read and what can we do to help them? In M. Grades, P. van den Broek, & B. Taylor (Eds.), *The first R: A right of all children* (pp 160–188). NY: Columbia University Press.
- Trabasso, T., van den Broek, P. & Suh, S. (1989). Logical necessity and transitivity of causal relations in the representation of stories. *Discourse Processes*, 12, 1–25.
- van den Broek, P. (1990). *The causal inference maker: Towards a process model of inference generation in text comprehension*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A model of the self-explanation effect. *Journal of the Learning Sciences*, 2, 1–60.
- Wolfe, B. W. & Goldman, S. R. (2005). Relations between adolescent's text processing and reasoning. *Cognition and Instruction*, 23, 467–502.
- Yuill, N. & Oakhill, J. (1988). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology*, 79, 173–186.

Part III

Metacomprehension

6 Metacognitive Monitoring During and After Reading

Keith W. Thiede

Boise State University

Thomas D. Griffin and Jennifer Wiley

University of Illinois at Chicago

Joshua S. Redford

Boise State University

Models of self-regulated learning (e.g., Butler & Winne, 1995; Metcalfe, 2002; Nelson & Narens, 1990; Thiede & Dunlosky, 1999) describe learning as a process that involves setting a learning goal, monitoring progress toward that goal, and regulating study in order to achieve that goal. For example, consider a student preparing for an upcoming test. As the student studies, she monitors her progress toward the goal of mastering the material. If her monitoring indicates that she has not yet mastered the material, she will likely re-study the material until her monitoring suggests that the material has been mastered, at which time she will stop studying. Accurate metacognitive monitoring is critical to effective regulation of study (Thiede, 1999; Winne & Perry, 2000). If someone does not accurately differentiate well-learned material from less-learned material, they could waste time studying material that is already well learned or fail to re-study material that has not yet been adequately learned.

Thiede, Anderson, and Therriault (2003) demonstrated the important role that accurate monitoring plays in learning from text by manipulating monitoring accuracy and examining the effect on regulation of study and learning outcomes. In this experiment, an initial important finding was that participants who were given a delayed-keyword-generation instruction (i.e., asked to generate a list of five keywords that captured the essence of a text prior to judging comprehension) more accurately monitored their comprehension on a first set of tests than did participants in a no-keyword-generation or immediate-keyword-generation group. Next, all participants were given the opportunity to select texts for re-study. The participants in the delayed-keyword group made better decisions about which texts to re-read than did participants in the other two groups. The mean proportion correct on first tests for the texts selected for re-reading versus those not selected for re-reading was 0.27 versus 0.78 respectively. The large difference in performance between these sets of texts indicates that this group was able to accurately discriminate the texts they comprehended well, from those they did not, and specifically chose those texts with poor comprehension for re-study. By contrast, for the other two groups, the mean proportion correct on first tests for the texts selected for re-reading versus those not selected for re-reading was 0.43 versus 0.53, respectively. These participants were either less able to discriminate what they had learned well than the delayed-keyword group, or were less likely to use their discrimination as a basis for re-study selections. More important, the more effective regulation of study observed in the delayed-keyword group led to higher overall reading comprehension performance as assessed by a second set of tests. Thus, this study demonstrated that improving

metacomprehension accuracy improved self-regulated study, and ultimately improved comprehension.

Improving Monitoring Accuracy

The majority of investigations exploring factors that affect the accuracy of metacognitive monitoring have been done using associative learning tasks (e.g., learning word pairs, translations or definitions). This research has shown that relative monitoring accuracy, as operationalized by computing intra-individual correlations (usually gamma correlations) between judgments of learning and test performance across items, improves when a person monitors learning (a) after a delay rather than immediately after studying an item (Dunlosky & Nelson, 1992, 1997; Kelemen & Weaver, 1997; Nelson & Dunlosky, 1991; Weaver & Kelemen, 1997), (b) after practice monitoring his or her learning (Vesonder & Voss, 1985), and (c) following a practice test of the material (King, Zechmeister, & Shaughnessy, 1980; Lovelace, 1984; Shaughnessy & Zechmeister, 1992). Furthermore, under certain circumstances, people can attain near perfect levels of monitoring accuracy. For example, monitoring accuracy was +0.93 for delayed judgments of learning in Nelson and Dunlosky (1991).

In sharp contrast to accuracy of monitoring during associative learning tasks, metacomprehension accuracy has typically been quite low. Standard metacomprehension paradigms use measures of predictive, relative accuracy from expository texts. Similar to metamemory approaches, readers are asked to read a set of texts, make judgments of learning or predictions about their performance on future tests, and then complete the tests. Again, intra-individual correlations (traditionally gamma correlations) are computed from this, and the average hovers around 0.27. Maki (1998a) reported this figure as the mean intra-individual correlation between comprehension ratings and test performance across 25 studies from her lab. A recent analysis by Dunlosky across his lab's studies arrived at the exact same figure (Dunlosky & Lipko, 2007). Other reviews by Lin and Zabrocky (1998) and Weaver, Bryant, and Burns (1995) reached similar conclusions.

Factors Contributing to Poor Monitoring Accuracy when Learning from Text

Monitoring learning from text may be a task constrained by rather different factors than monitoring learning of paired-associates (Wiley, Griffin & Thiede, 2005). Consider the task of judging learning in an associative learning task. A person studies a list of word pairs (e.g., dog-spoon) and is instructed that the test will involve recalling the second word when given the first. After studying the list of word pairs, the person is given the first word and asked to judge how confident he is that he will recall the second word in the future. In this case, the person should be perfectly clear about what he is being asked to judge, and this expectation maps perfectly onto the test. As noted above, delayed judgments of learning (JOLs) can produce nearly perfect monitoring accuracy, particularly when the prompt for judgments is the stimulus-alone cue (cf. Dunlosky & Nelson, 1992, who showed a large drop in accuracy when the judgment is prompted with a stimulus-response pair).

Contrast this with the process of judging learning from text. In the original Glenberg and Epstein (1985) paradigm, a person read a series of 16 200–400 word texts. The person then rated his comprehension for each text on a 7-point scale, and answered one inference question per text to provide a measure of comprehension. Metacomprehension accuracy was computed for each individual by correlating judgments of comprehension and test performance across the texts.

As noted above, such accuracy is generally observed to be quite poor, and several potential reasons can be gleaned by noting differences between learning from text and paired associates paradigms. These differences largely stem from the inherently greater complexity and vagueness of what it means to comprehend a text versus what it means to recall a word when cued by its associate. In the remainder of this chapter, we will review research that focuses on several obstacles to effective monitoring while learning from texts and potential solutions that have been offered to address the complexity and vagueness of the monitoring task. First, we briefly describe these constraints, then we explore them more fully in light of interventions designed to overcome them.

Some of the blame for poorer accuracy and null or unreliable findings in metacomprehension compared to metamemory research probably lies with the lower validity in the performance measures of text comprehension versus cued word recall. Actually, research on “monitoring accuracy” does not solely measure monitoring accuracy, but rather measures judgment-performance covariance, which is just as influenced by the validity of the performance measures as by the accuracy of learners’ judgments. The validity of our inferences about monitoring accuracy depends on the validity of the performance measures that greatly constrain any judgment-performance covariance. Metamemory research largely avoids this problem, because a valid recall measure requires little more than assessing recall for each word pair that was judged. However, texts and the comprehension of them are so complex and multi-dimensional that it is difficult to adequately define the construct, let alone measure it for a particular text. One implication is that we must carefully consider variability in the validity of our performance measures before drawing too many inferences about changes or differences in monitoring accuracy between conditions, studies, or paradigms. We should also take steps to improve the validity of our comprehension measures.

With comprehension measures, a major validity issue is the one of content coverage. Weaver (1990) discussed the problem that a one-item test does not provide a valid or reliable measure of comprehension, and demonstrated that metacomprehension accuracy improves when multiple comprehension items are given at test. The failure to use comprehension tests with multiple items per text is one likely reason for the low correlations observed in several early studies in the metacomprehension literature (i.e., Glenberg & Epstein, 1985; Glenberg, Sanocki, Epstein, & Morris, 1987, Experiment 1).

Beyond issues of reliability, Wiley et al. (2005) expanded on the need for multiple comprehension items per text by arguing that any test that lacks complete coverage of the text information will lack validity and tend to under-estimate true monitoring accuracy levels. What counts as coverage depends on the units of text being judged. Whether readers are asked to judge a whole text or certain portions, there are components and their inter-relations that contribute to the unit of the text being judged. Comprehension of one component may correlate only modestly with other components. Thus, coverage of all the various components and their relations within a unit of text is a minimal requirement for the validity of a measure intended to assess comprehension of a text unit.

Without complete coverage, judgment-performance alignment will not only be unreliable but will systematically under-estimate true monitoring accuracy. For instance, imagine a person who accurately knows that he understood 50% of a text, so he gives a mid-range judgment. If the test items only constitute 50% coverage, then his performance-judgment alignment will be misleadingly poor, regardless of whether the test taps the 50% he knew (actual performance being higher than judged), or the 50% he did not know (actual performance being lower than judged). The only situation where incomplete coverage would overestimate a reader’s accuracy is the low-probability scenario where

the reader bases their judgment on only a portion of the text and that happens to be the same portion tapped by the test items. Performance measure validity is a problem to be considered and overcome, but there are other factors that constrain the actual rather than just computed accuracy of learners' judgments.

Dunlosky and Lipko (2007) address a different form of coverage problem. They argue that poor metacomprehension accuracy is a function of judgments that cover too much text material. They point out that grain size of the typical monitoring judgment (which is global and covering the whole text) is much larger than the grain size of the test items (which may concern specific concepts from within the text). Readers may struggle to compute a judgment covering many components of the text. This is not a problem in typical paired-associate tasks, because the judgment and performance grain sizes are identical and at the level of associated pairs.

Yet another factor that may contribute to poor monitoring accuracy when learning from text is a poor alignment between which cues a person uses for metacomprehension judgments and which cues predict performance on a test of comprehension. The notion that assessments can vary in the level of representation they tap is derived from Kintsch's theory of comprehension, which posits that text is mentally represented in multiple forms from surface memory of the words to a logically integrated conceptual model (Kintsch, 1994). For instance, a mismatch can occur if a person judges comprehension based on their ability to recall the words from the text, but the test assesses whether they can apply the concepts to a new scenario (Rawson, Dunlosky, & Thiede, 2000; Wiley et al., 2005). Again, paired-associate metamemory tasks sidestep this problem. For the highly accurate delayed-JOLs, the judgment prompt is identical to the test prompt; thus, there is no possible misalignment. Even when the judgment and test prompts differ, paired-associate learners still need to only monitor their memory for the words and not a potentially orthogonal level of conceptual understanding.

A final factor may be the differential demands of the metacomprehension versus the metamemory monitoring process. Metacomprehension monitoring, more so than metamemory monitoring, may require concurrent attention at two levels of processing: directing attention to the monitoring of cues plus directing attention to processing the text. The inherent complexity of text processing and the comprehension process seem to entail that most valid cues will only be available at the time of judgment if they were attended to during reading itself. Based on this assumption, poor metacomprehension accuracy may occur especially among low-ability readers due to high concurrent processing requirements (Griffin, Wiley, & Thiede, 2008). Yet again, delayed-JOL metamemory tasks do not pose this obstacle when the judgment prompt is identical to the test prompt, because the judgment prompt itself gives rise to many if not most cues that learners need to predict future test performance.

To provide a current overview of work in the metacomprehension literature, Table 6.1 presents all studies that have been done in the area. The table is restricted to published experimental papers that use a predictive, relative accuracy paradigm with expository texts. For each study, the mean intra-individual correlation (γ) for each condition is reported. As noted above, due to issues of reliability of tests with only 1 or 2 items (Weaver, 1990) only studies with more than two items per test are included. Interestingly, an average of the "standard" predictive accuracy conditions in this table also leads to an average γ of 0.27. However, there are also several recent studies that have produced substantially higher levels of accuracy, and these studies form their own distribution toward the top end of the range. The second half of this chapter will describe how several lines of recent research have provided evidence for the various constraints discussed above, and how they may be overcome to improve monitoring accuracy for texts to substantially higher levels.

Table 6.1 A Review of the Literature on Metacomprehension Accuracy

<i>Experimenters</i>	<i>Texts</i>	<i>Judgment</i>	<i>Test</i>	<i>N</i>	<i>Condition</i>	<i>Predict G</i>	<i>Postdict G</i>	<i>Predict G Effect Size</i>
Anderson & Thiede, 2007	5 (~800 wds)	JOC 6pt	10 MC (inf)	87	No Summary	.20		.51 (summary conds.)
	5 (~800 wds)	JOC 6pt	10 MC (inf)	(87)	Immediate Summary	.31		
	5 (~800 wds)	JOC 6pt	10 MC (inf)	(87)	Delayed Summary	.64		
Baker & Dunlosky, 2006, exp. 1	8 (~400 wds)	PP 0-100	8 MC (4 det, 4 inf)	30	Immediate Judgment	.06		
				32	Delayed Judgment	.02		
exp. 2	8 (~400 wds)	PP 0-100	8 MC (4 det, 4 inf)	33	Recall before Delayed Judgment	.10		
				31	Recall after Delayed Judgment	.18		
Dunlosky, Baker, & Rawson, 2006	8 (~400 wds)	PP 0-100	8 MC (4 det, 4 inf)	33	Younger Adults	.16		
				22	Older Adults	.23		
Dunlosky & Rawson, 2005	6 (~400 wds)	PP 0-100	6 MC (3 det, 3 inf)	39	Read once	.29		
				38	Immediate re-reading	.53		
Dunlosky, Rawson, & Hacker, 2002	6 (~400 wds)	PP 0-100	not stated?	36	Delayed re-reading	.22		
			6 MC (3 det, 3 inf)	38	No advance organizer	.36		
Dunlosky, Rawson, & McDonald, 2002, exp. 2	6 (~300 wds)	PP 0-100	4 cued recall	35	Advance organizer	.42		
				38	Global prediction	.37		
Dunlosky, Rawson, & Middleton, 2005, exp. 1	6 (~300 wds)	PP 0-100	4 cued recall	(38)	Term-specific prediction	.38	.72	
				26	Global Immed Judgment	.52		
exp. 2				(26)	Term Immed Judgment	.57		
				27	Global Delayed Judgment	.05		
				(27)	Term Delayed Judgment	.57		
				24	Recall before Global Immed	.41		
				(24)	Recall before Term Immed	.73		
				26	Recall before Global Delayed	.52		
				(26)	Recall before Term Delayed	.64		
		6 (~300 wds)	PP 0-100	4 cued recall	39	Global Immed Judgment	.23	
					39	Term Immed Judgment	.61	
					40	Scored Recall before Global	.34	
				40	Scored Recall before Term	.57		

(Continued overleaf)

Table 6.1 Continued

Experimenters	Texts	Judgment	Test	N	Condition	Predict G	Postdict G	Predict G Effect Size
Griffin, Wiley, & Thiede, 2008, exp. 1 exp. 2	5 (~800 wds)	JOC 7-pt	~18 TF inf	36	Read once	.27		.337
				36	Reread	.45		
	5 (~800 wds)	JOC 7-pt	~18 TF inf	30	Read once	.22		.212 (read vs. reread)
				30	Reread	.35		.461 (reread vs. self-explain)
				30	Reread and Self-explain	.67		
Jee, Wiley, & Griffin, 2006	5 (~400 wds)	PP 6 pt	5 short answer	98	Multiple Texts on Baseball Topics	.20		
Lefevre & Lories, 2004 exp. 1 exp. 2	6 (~550 wds)	JOC 0-20 pt	8 short answer	56	Immediate/paragraph judgments	.36		
				(56)	Delayed/global judgments	.43		
	6 (~550 wds)	JOC 0-20 pt	8 short answer	68	Immediate/paragraph judgments	.27		
			(68)	Delayed/global judgments	.38			
Lin, Moore, & Zabrocky, 2001	12 (~200 wds)	CONF 7-pt	4 TF inf	60	Read once	.14	.28	
Lin, Zabrocky, & Moore, 2002	2 exp./2 narr.	CONF 6-pt	16 MC (8 det/them)	120*	Easy	.19		.398 *(60 old/60 young)
	2 exp./2 narr.	CONF 6-pt	16 MC (8 det/them)	(120)	Standard	.47		
	2 exp./2 narr.	CONF 6-pt	16 MC (8 det/them)	(120)	Difficult	-.09		
Maki, 1995 exp. 1 exp. 2	12 (~200 wds)	PP 6-pt	3 MC (det)	18	Immed. reread, Important ideas	.15	.56	
			3 MC (det)	18	Immed. reread, Unimportant ideas	.01	.70	
			3 MC (inf)	18	Immed. reread, Higher ideas	.27	.22	
	12 (~200 wds)	PP 6-pt	3 MC (det)	18	RSVP, Important ideas	.00	.22	
			3 MC (det)	18	RSVP, Unimportant ideas	.43	.47	
			3 MC (inf)	18	RSVP, Higher ideas	.36	.34	
Maki, 1998 exp. 1 exp. 2	4 (~400 wds)	PP	16 MC (det)	24	Multiple Texts/Topics	.28		
	1 (~1600 wds)	PP	64 MC (det)	24	Single Text/Topic	.14		
	6 (~400 wds)	PP	5 MC	32	Familiarity predicts Performance	.38		
	6 (~400 wds)	PP	5 MC	32	Familiarity Does not Predict	.37		
Maki, 1998	6 (~400 wds)	PP 6-pt	5 MC	24	Immediate Rating and Test	.47	.62	.323
				24	Immed. Rating, Delayed Test	.05	.40	
				24	Delayed Rating, Immed. Test	.20	.49	
				24	Delayed Rating, Test	.02	.46	

Maki, 1998, exp. 1	6 (~400 wds)	PP 6-pt	10 MC	16	Low Expect, Poorly anchor	.34	.47
				16	Low Expect, No anchor	.03	.45
				16	Low Expect, Well anchor	.04	.36
				16	High Expect, Poorly anchor	.27	.47
				16	High Expect, No anchor	.18	.33
				16	High Expect, Well anchor	.47	.63
exp. 2	6 (~400 wds)	PP 6-pt	10 MC	23	Low Expect, Poorly anchor	.13	.30
				24	Low Expect, Well anchor	.14	.19
				23	No Expect, Poorly anchor	.23	.36
				24	No Expect, Well anchor	.17	.42
				23	High Expect, Poorly anchor	.12	.39
				24	High Expect, Well anchor	.13	.57
Maki, Foley, Kajer, Thompson, & Wilert, 1990, exp. 1	1 (21–27 para)	EOC 6-pt	2 cued recall per p	40	EOC-intact text	.26	.45
				40	EOC-deleted letters	.13	.55
		CONF 6-pt		(40)	CONF-intact text	.16	.56
				(40)	CONF-deleted letters	.31	.59
exp. 2	1 (~14–17 para)	CONF 6-pt	2 cued recall per p	40	intact text	.18	.52
				40	deleted letters	.37	.53
Maki, Jonas, & Kallod, 1994	1 (~1700 wds)	PP 6-pt	4 MC	75	read once	.11	.55
Maki & Serra, 1992a exp. 1	12 (~200 wds)	PP rank titles	4 MC	36	practice test	.22	.35
				36	no practice test	.23	.32
exp. 2	12 (~200 wds)	PP rank titles	3 MC	18	identical to practice	.32	.24
				18	stem from practice	.23	.24
				18	similar to practice	.10	.45
				18	no practice	.27	.40
exp. 3	12 (~200 wds)	PP rank titles	3 MC	18	no feedback on similar practice test	.29	.37
				18	feedback on similar practice test	.06	.21
Maki & Serra, 1992b	12 (~200 wds)	EOC rank titles	4 MC	48	comprehension judgment	.17	.28
exp. 1		PP rank titles		48	performance judgment	.31	.40
exp. 2	12 (~200 wds)	EOC rank titles	4 MC	24	comprehension judgment	.14	.34
		PP rank titles		24	performance judgment	.40	.41

(Continued overleaf)

Table 6.1 Continued

<i>Experimenters</i>	<i>Texts</i>	<i>Judgment</i>	<i>Test</i>	<i>N</i>	<i>Condition</i>	<i>Predict G</i>	<i>Postdict G</i>	<i>Predict G Effect Size</i>
exp. 3	12 (~200 wds)	EOC rank titles	6 MC	18	comprehension judgment	.11	.25	
Maki, Shields, Wheeler, & Zaccchilli, 2005	6 (~400 wds)	PP rank titles	6 MC (3 det, 3 inf)	18	performance judgment	.25	.31	
	6 (~400 wds)	PP 7 pt	6 MC (3 det, 3 inf)	52	Easy Texts	.20	.36	Not reported
	6 (~400 wds)	PP 7 pt	6 MC (3 det, 3 inf)	55	Mixed Set	.31	.42	
	6 (~400 wds)	PP 7 pt	6 MC (3 det, 3 inf)	51	Hard Texts	.36	.57	
Miesner & Maki, 2007	6 (~400 wds)	PP 0-100	5 MC	38	Low Test Anxiety	.20	.19	
				42	High Test Anxiety	.58	.46	
Olin & Zelinsky, 1997	4 (9 para)	JOC 6 pt	36 TF sent recog	51	Younger	.11		
				52	Older	.13		
Pierce & Smith, 2001, exp. 2	3 (~500 wds)	PP 5 pt	16 MC (8 det/them)	47	read once	.28	.53	
Rawson & Dunlosky, 2007	6 (~300 wds)	PP 0-100	4 cued recall	26	Term recall/score without answer	.68		
				30	Term recall/score with answer	.52		
				(26)	Global/score without answer	.21		
				(30)	Global/score with answer	.36		
Rawson, Dunlosky, & Thiede, 2000, exp. 1	6 (~400 wds)	PP 0-100	6 MC (inf?)	40	read once	.24	.50	.298
				40	rereading	.57	.47	.413
exp. 2	6 (~200 wds)	JOC 7-pt	6 MC (inf?)	30	read once	.19		
				30	rereading	.55		
Thiede & Anderson, 2003, exp. 1	6 (~200 wds)	JOC 7-pt	6 MC (thematic?)	25	No Summary	.28		.308
				25	Immediate Summary	.24		
				25	Delayed Summary	.60		
exp. 2	6 (~1300 wds)	JOC 7-pt	12-MC (6 det, 6 inf)	30	No Summary	.28		.262
				30	Immediate Summary	.35		
				30	Delayed Summary	.61		
Thiede, Anderson, & Theriault, 2003	6 (~1300 wds)	JOC 7-pt	6-MC (3 det, 3 inf)	22	No keyword generation	.38		.351
				22	Immediate keyword generation	.29		
				22	Delayed keyword generation	.70		

Thiede, Dumlosky, Griffin, & Wiley, 2005, exp. 1	6 (~1300 wds)	JOC 7-pt	12-MC (6 det, 6 inf)	33	Delayed Keyword and Judgment	.52	.315 (delay vs. immediate) Not significant (lag vs. no-lag)	
					Delayed Keyword, Immed. Judge	.55		
					Immediate Keyword, Delay Judge	.27		
					Immediate Keyword and Judge	.29		
exp. 2	6 (~1300 wds)	JOC 7-pt	6 MC (6 inf)	25	Delayed, Successive Keyword	.57	.354 (delay vs. immediate)	
					Delayed Spaced Keyword	.55		
					Immediate Spaced Keyword	.16		
					Short delay, Spaced Keyword	.39		
exp. 3	6 (~1300 wds)	JOC 7-pt	6 MC (6 inf)	25	Immediate-think-about-text	.25	Not significant	
					Delayed-think-about-text	.27		
					Delayed Keyword generation	.52		
exp. 4	6 (~1300 wds)	JOC 7-pt	6 MC (6 inf)	33	Delayed keyword reading	.27	.338	
					33			
Thiede, Griffin, Wiley, & Anderson, in press, exp. 1	5 (~800 wds)	JOC 7-pt	10 MC (inf)	74	Normal reader, No Summary	.21	.276 (normal vs. at-risk) .196 (summary conds.)	
					(74)	Normal reader, Immed Summary		.29
					(74)	Normal reader, Delayed Summary		.63
					32	At risk reader, No Summary		.08
					(32)	At risk reader, Immed Summary		.17
					(32)	At risk reader, Delayed Summary		.48
					21	At risk, Immediate Judgment		.30
					(21)	At risk, Delayed Judgment		.33
exp. 2	5 (~300 wds)	JOC 7-pt	5 MC (inf)	(21)	At risk, Concept Map Generation	.66	.424 (conds.)	
					(21)			
Thomas & McDaniel, 2006, exp. 1	6 (~300 wds)	PP 0-100 scale	6 cued recall (det)	39	letter reinsertion, detail	.41	.334 (task) .489 (task x test)	
					6 cued recall (them)	-.33		
					6 cued recall (det)	-.06		
					6 cued recall (them)	.59		
					6 cued recall (det)	.27		
					6 cued recall (them)	.11		
					6 cued recall (det)	.62		
					6 cued recall (them)	-.13		
					6 cued recall (det)	.06		
					6 cued recall (them)	.55		
exp. 2	6 (~300 wds)	PP 0-100 scale	6 cued recall (det)	25	one item test	.09	.494 (task x test) Not reported (task)	
					6 cued recall (them)	.06		
					6 cued recall (det)	.55		
					6 cued recall (them)	.55		
Weaver, 1990, exp. 1 exp. 2	16 (~200 wds)	PP 6-pt	1 TF (inf)	21	one item test	.09	Not reported Not reported	
					1 TF (inf)	.02		
					2 TF (inf)	.28		
					4 TF (inf)	.40		

(Continued overleaf)

Table 6.1 Continued

Experimenters	Texts	Judgment	Test	N	Condition	Predict G	Postdict G	Predict G Effect Size
Weaver & Bryant, 1995 exp. 1	4 (~500 wds)	JOC 6-pt	8 MC Them	49 (49)	Thematic Detailed	.22		.268
			8 MC Detail			.39		
exp. 2	2 narr/2 exp	JOC 6-pt	8 MC Them	30 (30)	Easy-Thematic Easy-Detailed	.20		.267
			8 MC Detail			.38		
exp. 3	2 narr/2 exp	JOC 6-pt	8 MC Them	31 (31)	Standard-Thematic Standard-Detailed	.75		
			8 MC Detail			.63		
			8 MC Them			.33		
			8 MC Detail			.30		
			8 MC Them			.15		
			8 MC Detail			.37		
Wiley, Griffin & Thiede, 2008, exp. 1	2 narr/2 exp	JOC 6-pt	8 MC Them	31 (31)	Standard-Thematic Standard-Detailed	.72		
			8 MC Detail			.67		
			8 MC Them			-.04		
			8 MC Detail			.45		
			5 MC Detail			.23		
			5 MC Inference			.22		
exp. 2	6 (~800 wds)	PP 6-pt	5 MC Detail	36 (36)	No Test Expectancy No Test Expectancy Expect Memory Test Expect Memory Test Expect Inference Test Expect Inference Test No Expectancy No Expectancy Expect Inference Test Expect Inference Test Self Explain Self Explain Expect Inference, Self Explain Expect Inference, Self Explain	.38		
			5 MC Inference			.18		
			5 MC Detail			.17		
			5 MC Inference			.36		
			5 MC Detail			.34		
			5 MC Inference			.16		
			5 MC Detail			.28		
			5 MC Inference			.38		
			5 MC Detail			.15		
			5 MC Inference			.32		
			5 MC Detail			.20		
			5 MC Inference			.51		

PP = performance prediction, e.g., “How well do you think you will be able to answer test questions over this material?” 0 (definitely won't be able), . . . , 100 (definitely will be able) or number of items
 JOC = judgment of comprehension, e.g., “How well do you think you understand the text you just read?” 1 (not at all) to 7 (very well)
 EOC = judgment of how easy it was to comprehend a text
 CONF = judgment of confidence, e.g., “report your confidence that you will be able to use what you learned to answer test questions” 1 (very low) to 6 (very high)
 exp. = experiment, MC = multiple choice, TF = true/false, det = detail questions, inf = inference questions, them = thematic questions, G = gamma correlations
 All texts are expository (exp) unless otherwise noted as narr = narrative
 N in parenthesis indicates repeated measures design

Aligning the Grain-size of Judgments and Tests: Term-specific Judgments

Metacomprehension judgments usually ask individuals to make a global assessment of their text comprehension—a general evaluation of their overall understanding of the material. Although individuals may be sensitive to their level of comprehension for specific concepts, it is unclear whether they would be able to accurately translate this into a single value. Moreover, criterion tests often involve testing people's learning of specific concepts within a passage, which also may not map cleanly onto a global judgment. Thus, the use of a single global judgment and the mismatch in grain size between monitoring judgments and subsequent tests provides one potential explanation for the low monitoring accuracy reported in previous research (Dunlosky & Lipko, 2007).

Dunlosky, Rawson, and Middleton (2005) tested the possibility that a mismatch in grain size undermines accurate measurements of monitoring accuracy while learning from text by reducing the grain size of the monitoring judgments. In addition to asking participants to make a global judgment of their level of comprehension, participants also made monitoring judgments regarding their ability to recall the meaning of specific terms contained in the text. Surprisingly, the accuracy of these term-specific memory judgments was not dramatically different than the accuracy of the global comprehension judgment.

Dunlosky et al. (2005) discussed two issues that might prevent a concordance in grain size from improving monitoring accuracy when learning from text. First, participants were relying on their familiarity with specific terms rather than attempting to recall term definitions prior to making their judgments. When forced to attempt recall of definitions prior to making judgments, accuracy increased to 0.73 (see also Dunlosky, Rawson, & McDonald, 2002, for a similar result). Second, in agreement with Koriat's accessibility model (1993), participants who attempted recall prior to their judgments partly based their judgments on the overall quantity of retrieved information rather than on the quality of the retrieved information. In a subsequent study, Rawson and Dunlosky (2007) found that participants who received feedback regarding the accuracy of their recalled definitions (the original definition from the text being displayed next to their typed definition) prior to their judgments further improved their monitoring accuracy to 0.92.

Note that the grain-size approach also addresses the content coverage issue discussed above. Thus, another way that grain size may be helping the concordance between predicted performance and actual performance on the test is by making the coverage of the test obvious. As we discuss below, the nature of the test (in this case, a test of memory for specific definitions) is also made explicitly clear to learners through this procedure. A final way that this procedure may aid accuracy is through providing the exact test items, and giving students the correct answers to the exact test items. With this manipulation, the judgment that is made changes from a prediction to a postdiction, and thus it becomes more similar to the conditions that promote high levels of performance in paired-associate paradigms, and replicates the robust effects that have been found across many studies that postdictions are superior to predictions of metacomprehension accuracy (Glenberg & Epstein, 1985, 1987; Glenberg, Sanocki, Epstein, & Morris, 1987; Lin, Moore, & Zabrocky, 2001; Maki, Foley, Kajer, Thompson, & Willert, 1990; Maki & Serra, 1992a, 1992b; Pierce & Smith, 2001). Nevertheless, this research demonstrates that a mismatch in grain size can produce lower levels of monitoring accuracy when learning from text. When grain size is matched and other risk factors are accounted for, monitoring accuracy can increase dramatically, and the present levels are well beyond the magnitude of the improvements that are typically seen as a result of exposure to the test items.

Aligning the Level of Understanding Assessed by Judgments and Tests

Another factor undermining monitoring accuracy when learning from text pertains to the perspective individuals adopt when reading a text and assessing their level of text comprehension (Wiley et al., 2005). Most metacomprehension research has employed expository texts and asked students to judge their level of comprehension without further specification of what the construct of comprehension might mean. Although meaningful comprehension of expository texts requires an appreciation of the information at a deeper level (e.g., how or why some event occurs), many students approach learning from texts in the same way as they approach learning word lists—as a collection of discrete pieces of information to be stored for later retrieval. To the extent that comprehension tests are more concerned with a reader's ability to make inferences, connections or develop an understanding of how or why some event or phenomenon occurs, then metacomprehension judgments may be inaccurate because individuals are basing their judgments on cues (e.g., memory of details) that are not predictive of test performance (i.e., the ability to make inferences). Thus, a second explanation for the low monitoring accuracy reported in previous research is the mismatch between the kinds of cues that readers use to predict their own comprehension, and the kinds of cues that they should use.

When assessing comprehension, a person can monitor various cues that are produced by comprehension processes, such as the fluency of text processing (Dunlosky & Rawson, 2005; Rawson & Dunlosky, 2002; Lefevre & Lories, 2004; Maki, Shields, Wheeler, & Zacchilli, 2005; Rawson, Dunlosky, & McDonald, 2002) or their familiarity with the topic (Maki, 1998a; Jee, Wiley, & Griffin, 2006) or ability to remember parts of the text (Thiede, Griffin, Wiley, & Anderson, in press). A judgment of comprehension is then based on an inference about those cues (Schwartz, Benjamin, & Bjork, 1998). In the case of fluency, people presumably judge that their understanding of a text is better when the text is easily processed than when it is difficult to process (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989). Similarly, people may judge that their understanding of a text is better when elements of the text are easily recalled, or when they are interested in or familiar with the topic. According to this cue-utilization framework of metacognitive monitoring (Koriat, 1997), the accuracy of comprehension judgments will be a function of the degree to which the cues used in assessing comprehension empirically correlate with performance on comprehension tests. More specifically, assuming the inferences are valid, metacomprehension accuracy will increase as the available cues more highly correlate with subsequent test performance. The particular problem in the case of comprehension judgments is that readers often do use fluency, topic, and memory-based cues to judge their understanding, but these cues are not necessarily predictive of actual comprehension (Thiede et al., in press). To determine the types of cues that are highly predictive of performance on tests of comprehension, it is necessary to understand the comprehension processes that provide many of those cues. According to the construction-integration model of comprehension (Kintsch, 1998), readers construct meaning from text at several levels of representation: a lexical or surface level, a textbase level, and a situation model level. The lexical level, containing the surface features of the text, is constructed as the words and phrases appearing in the text are encoded. The textbase level is constructed as segments of the surface text are parsed into propositions, and as links between text propositions are formed based on argument overlap and other text-explicit factors. However, the deeper understanding of the text, and the scenario it describes, is constructed at the level of the situation model, which involves connecting text information with the reader's prior knowledge and using it to generate inferences and implications from the text. It is a

reader's situation model that largely determines his or her performance on tests of comprehension (McNamara, Kintsch, Songer, & Kintsch, 1996). Therefore, getting people to base their judgments on cues related to their situation model rather than their surface model or textbase should increase the predictive accuracy of judgments when tests are comprehension-based (Rawson et al., 2000; Wiley et al., 2005).

This situation model approach has been supported by a number of studies that have improved metacomprehension accuracy through manipulations designed to prompt readers to access and utilize their situation models during or after reading. These manipulations have included generating summaries or keywords after a delay (Anderson & Thiede, 2008; Thiede & Anderson, 2003; Thiede et al., 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005), engaging in self-explanation while reading (Griffin et al., 2008; Wiley, Griffin, & Thiede, 2008), and engaging in concept mapping while reading (Thiede et al., in press).

Delayed-Generation Effect

In several studies, Thiede and his colleagues have found that having students engage in delayed-generation tasks produced unprecedented levels of metacomprehension accuracy (Anderson & Thiede, 2008; Thiede & Anderson, 2003; Thiede et al., 2003; Thiede et al., 2005). In Thiede and Anderson (2003), metacomprehension accuracy was dramatically greater for a delayed-summary group (mean correlation around 0.60) than for other groups (mean correlation around 0.26) across two experiments (one with shorter texts and one with longer texts).

Thiede et al. (2003) showed that generating a list of five keywords that captured the essence of a text produced a similar boost in metacomprehension accuracy (mean correlation of 0.71) as had generating summaries. Moreover, the effect was again influenced by the timing of generation. That is, generating keywords after a delay produced dramatic improvement, whereas generating keywords immediately after reading did not affect accuracy compared to a control group (mean correlation of both around 0.25).

As noted above, the cue-utilization framework of metacognitive monitoring (Koriat, 1997) and the construction-integration model of comprehension (Kintsch, 1998) provide a framework to explain why delayed-generation tasks improve metacomprehension accuracy.

Writing a summary or generating keywords may allow a reader to reflect on how successfully he or she can retrieve information during the generation task (cf. the modified feedback hypothesis described by Glenberg et al., 1987). Accordingly, a text may receive a high rating of comprehension if the person was able to retrieve a great deal of information about the text during the generation task; whereas, a text may receive a low rating of comprehension if the person struggled to retrieve information about the text during summarization or keyword generation.

Further, the timing of the generation task is critical. Kintsch, Welsch, Schmalhofer, and Zimny (1990) showed that memory for surface and textbase representations of text decays over time, whereas the situation model is robust to such decay. Thus, when writing a summary or generating keywords immediately after reading, a person may have easy access to the lexical and textbase representations and the person can use this information for generation. However, feedback gained during immediate generation tasks may not provide an accurate basis for judgments of comprehension because performance on the generation tasks is being determined largely by the surface and textbase representations. In contrast, when writing a summary after a delay, the findings by Kintsch et al. (1990) suggest that a person will likely have access primarily to the situation model for the text.

Thus, using delayed summary and keyword generation tasks provides more valid cues for judgment of comprehension because both the comprehension judgments and test performance are based on the situation model.

A stronger test of the situation model approach was to determine which delays might be responsible for the delayed-generation task effect observed in previous studies (Thiede & Anderson, 2003; Thiede et al., 2003). According to the situation model approach, the critical delay for better metacomprehension should be the one between reading and summarizing or keyword generation because this delay causes readers to access their situation model prior to judgment. The generation task forces readers to access and use their text representation, and doing so at a delay after reading means that their representation is more likely to be based on a situation model than the surface or textbase level because they should have decayed during the delay. Although this interpretation fits the findings of Thiede and Anderson (2003) and Thiede et al. (2003), there were actually several features of the previous studies' designs and procedures that were confounded and preclude them from providing clear support for the situation model approach. In Thiede et al. (2003), the delayed condition was created by having readers generate keywords for all texts after reading all the texts, rather than immediately after each text. This produced not only a delay between reading and keyword generation, but also decreased the time lag between keyword generation and judgments of comprehension, and decreased the time lag between each keyword generation by having all generation tasks performed in succession for the delayed group. In contrast to the reading-keyword delay, these other two time lags could account for the superior accuracy in the delay group without implying anything about access to the situation model.

To test for the viability of this alternative explanation, Thiede et al. (2005) manipulated the possible effects of different lags on metacomprehension accuracy. As altering the order of tasks in the protocol will always affect more than one time lag, a set of two experiments was constructed, where each one evaluated the possible effects of one of the confounding lags. Both experiments replicated the earlier findings and showed that the delay between reading and generating keywords was critical for improving metacomprehension accuracy. Experiment 1 also showed that the keyword-judgment lag was not critical, while Experiment 2 showed that the lag between multiple keyword tasks was not critical. In addition, Experiments 3 and 4 showed that the critical factor was actually generating keywords at a delay and not merely thinking about the text or being re-exposed to key ideas of the texts at a delay.

Thus, this set of four experiments provided a compelling case that the act of accessing and making use of one's representation via a generation task at a delay after reading are the critical factors that have produced some of the largest improvements in metacomprehension accuracy observed to date. Moreover, these findings provide strong support for the situation model approach. They suggest that it is access to valid cues (i.e., those based in a situation model) that underlies why both the reading-keyword delay and the generation component of the keyword task combine to produce the observed improvements in metacomprehension accuracy.

Self-Explanation and Cue-Access

The previous findings suggest that getting readers to focus on their situation model during monitoring will improve metacomprehension accuracy. Another way to test this hypothesis is to examine interventions that have been used to get readers to construct more complete and coherent situation models. Such interventions should more generally increase readers' attention to their situation-model representations, giving them greater

access to cues about the quality of their situation model that they could use to more accurately predict their comprehension performance.

One such intervention that has improved situation-model construction is self-explanation. Instructing and/or training readers to explain the logical and causal connections among idea units while they read has been shown to improve the completeness and coherence of the mental models readers construct of complex causal phenomena (e.g., Chi, DeLeeuw, Chiu, & Lavancher, 1994; Hemmerich & Wiley, 2002; McNamara, 2004; Wiley & Voss, 1999; Wiley, 2001). The interventions have ranged from simple global task instructions (Wiley & Voss, 1999; Wiley, 2001) to targeted prompts to explain specific relations (e.g., Chi et al., 1994), to more extensive recursive training and feedback designed to improve the quality of self-explanations (e.g., McNamara, 2004). Based on these previous findings, a self-explanation instruction should prompt readers to focus on and use their situation models during reading. If this is the case, the situation model approach would also predict that self-explanation tasks should improve the accuracy of metacomprehension judgments during reading.

A recent study done by Griffin et al. (2008) has provided support for this hypothesis. In this study, giving readers a one page pre-reading instruction that prompted them to self-explain as they read the expository texts increased metacomprehension accuracy to levels on a par with those found for delayed-generation tasks (mean correlation around 0.60) as compared to 0.39 in control conditions. Wiley et al. (2008) recently found a similar result with a similar self-explanation condition. Another finding by Thomas and McDaniel (2007) may also be related, as they found improved monitoring accuracy specifically on a conceptual (rather than detail-oriented) test following a reading task that required the re-ordering of sentences (rather than inserting missing letters). This sentence-sorting manipulation could also have improved attention to the situation-model level.

A key difference between these studies and delayed-generation tasks used in previous studies is that the boost in accuracy occurred even without introducing a delay between reading and judging comprehension. Besides addressing potential pragmatic pedagogical concerns, this avoids an alternative explanation for the delayed-generation effects, namely a transfer-appropriate-monitoring account which would posit that increased accuracy is due to the fact that the processing during the generation task and the target test both occur at a delay. Instead, the self-explanation results provide specific support for the situation-model approach, that metacomprehension accuracy will improve when readers are made to focus on valid cues for predicting comprehension, rather than using cues based on familiarity, fluency, or surface memory for text.

Concept Mapping and Cue Access

An intervention similar to self-explanation that has been used in the literature, especially with younger or less-skilled readers, is concept mapping. Weinstein and Mayer (1986) suggested that instructing students to create concept maps of texts during reading helps them to identify the connections among concepts in a text. In discussion of concept mapping as an intervention, Weinstein and Mayer suggest that argumentation, self-explanation, and concept mapping tasks all help readers to construct and pay attention to the underlying causal, situation models of the subject matter. As the text is available during the activity, concept mapping tasks may be especially appropriate for less-able readers who may have difficulty engaging in a task that requires them to remember a text (Nesbit & Adesope, 2006; Stensvold & Wilson, 1990).

Using a within-subjects design, Thiede et al. (in press) gathered initial metacomprehension data by running participants through the standard experimental procedure for

assessing metacomprehension accuracy. That is, participants read a series of texts, judged their comprehension of each text, and then completed a comprehension test for each text. Participants, who were college students in a remedial reading course, then received eight days of instruction on how to construct concept maps for texts. They then completed the standard procedure for assessing monitoring accuracy again, but this time they constructed concept maps while reading the new set of texts. As with the self-explanation instructions, constructing concept maps while reading increased students' metacomprehension accuracy to around 0.65, which was a significant improvement from pre-intervention levels of around 0.32. These findings provide further support for the basic tenet of the situation-model approach that getting readers to access and use their situation model prior to judging comprehension improves their metacomprehension accuracy. And again, this boost in accuracy occurred without introducing a delay between reading and judgments.

The interventions discussed up to now have all dealt with increasing the alignment (of either grain size or level of representation) between judgments and performance measures as readers attempt to learn from text. We now turn to an accuracy constraint that deals with the demands of the monitoring task itself.

Demands of Monitoring during Reading

Unlike monitoring metamemory of paired associates, monitoring metacomprehension of text requires concurrent attention to two cognitive levels: processing the text and monitoring the products of comprehension. These levels have been defined previously as the object-level, where the incoming information such as a text is processed and a mental representation of it is formed, and the meta-level, where cues about one's mental representation are the information being processed and monitored (see Fischer & Mandl, 1984; Nelson & Narens, 1990). The simplicity of cued recall in associative learning paradigms means that the judgment prompt is largely identical to the testing prompt. Learners can simply perform the actual memory test at the time of judgment and use any performance feedback to make fairly accurate judgments about future test performance. They can attend to the meta-level after much of the object-level processing, the reading and studying of the word pair, is complete.

In contrast, text processing is more complex and involves many aspects of comprehension that typically are not (and cannot realistically be) fully contained in a judgment prompt. This is especially true of global judgments such as "How well did you understand that text?" Thus, valid monitoring cues for text comprehension judgments cannot be produced at the time of judgment, and must be generated during the reading of the text. Access to those cues during later judgment will depend on how much attention was paid to them during the processing of the text. Thus, accurate monitoring of text comprehension requires that readers concurrently attend to both processing the text at the object-level and processing cues about their level of comprehension at the meta-level during reading. (Although with extremely short or conceptually simple texts, most cues produced during reading may remain highly available even after reading.)

Griffin et al. (2008) tested the hypothesis that having to concurrently attend to (and/or repeatedly switch between) two processing levels limits the accuracy of metacomprehension judgments. In this study, judgment accuracy was shown to be predicted by individual differences in two abilities (e.g., comprehension ability and working memory capacity) that are theoretically related to the ability to monitor meta-level cues while concurrently processing the text during reading. Moreover, a re-reading manipulation demonstrated that these demands were alleviated and accuracy improved when readers were allowed to read the text twice prior to judgment. Benefits for re-reading were found

specifically for readers with poor comprehension ability, as well as for readers with low working memory capacity. This latter finding is especially telling as it implicates that poor monitoring is due to a more general inability to allocate or control attentional resources, which is the construct purportedly measured by working-memory span tasks (Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005). Presumably, readers lower in either working memory or comprehension ability struggled with concurrently attending to both the meta-level and object-level during a first reading, but could devote less attention towards text processing thus more towards meta-level cues and monitoring during the second reading.

This work differs from the rest reported in this chapter in that it deals with obstacles to accurate monitoring due to the demands of the monitoring process itself, which may exceed a reader's available resources, rather than obstacles due to mismatches in the nature of the cues that are used for judgment. It also extends other research that has demonstrated improvements in metacomprehension accuracy with re-reading manipulations (Rawson et al., 2000) and specifically for immediate rather than delayed re-reading (Dunlosky & Rawson, 2005). Although Rawson and colleagues argued that re-reading can make situation-model based cues more available, the fact that Griffin et al.'s (2008) re-reading effects were moderated by individual differences in attentional control, coupled with the fact that re-reading effects were modest compared to self-explanation effects (a manipulation that we presume does direct attention to the situation model) suggests that re-reading effects may depend more on freeing attentional resources than increasing the availability of situation-model based cues.

A final set of findings also seems to fit in with this proposed demand of monitoring constraint. The lone observation of predictive metacomprehension accuracy above 0.5 before the turn of the century was an interesting finding by Weaver and Bryant (1995) that high levels of metacomprehension accuracy can be observed when texts are moderately difficult (correlations between 0.66 and 0.69 across two experiments). Moderately difficult texts led to superior metacomprehension accuracy over either texts that were too easy, or texts that were too difficult in terms of their Flesch readability scores. A similar pattern using the same texts was replicated by Lin, Zabrocky and Moore (2002). In particular, the finding that moderately difficult texts lead to better metacomprehension than difficult texts is consistent with the demands of monitoring view because texts that are too difficult would leave no resources for comprehension monitoring. Explanations for the poor accuracy on easy texts have attributed poor performance to either a lack of discriminability among the easy set, or a lack of engagement (Dunlosky & Rawson, 2005; Weaver & Bryant, 1995).

Summary

Although traditionally levels of metacomprehension accuracy have been observed to be quite low, several studies in the last decade have demonstrated substantial increases in monitoring accuracy when readers are learning from text. The common thread that unites all of these attempts to improve monitoring during or after reading is that they address constraints that arise from the inherent complexity of monitoring learning from text. There are additional constraints and obstacles when learning involves text comprehension that are not faced when learning involves cued recall of paired associates. As such, this work may begin to provide insights into a puzzle that has faced metacognitive researchers for decades, namely why metacognitive accuracy suffers so greatly when the to-be-learned material goes from paired associates that must be recalled to texts that must be comprehended.

Most of the successful interventions presented here were based in a situation-model approach to improving metacomprehension accuracy. It is important to note that the situation-model approach does not predict that readers who form better situation models will necessarily have better metacomprehension (Griffin et al., 2008; Jee et al., 2006). Improving the quality of the representation that can be constructed is not the point of these interventions. Instead what is important is that readers access and use the situation model to inform their judgments of comprehension. Although self-explaining and concept mapping may also improve situation-model construction and may ultimately lead to better comprehension, the goal of these interventions was to direct readers to attend to valid cues for comprehension.

It is also important to distinguish the present situation-model approach from an earlier approach in much the same spirit: the levels-of-disruption hypothesis (Dunlosky & Rawson, 2005; Dunlosky, Rawson, & Hacker, 2002). The levels-of-disruption hypothesis shares the assumption that valid cues are to be found at the level of the reader's situation model. However, it posits that the valid cues are generated in response to disruptions that occur during the construction of a situation model. The present situation-model approach foregoes the additional assumption that feedback from disruptions provides the cues that should be used to predict comprehension. Although this is one possible cue that a reader might use, it might not actually predict comprehension all that well. Disruptions resulting from construction and integration of representations may only indicate how difficult the process of comprehension was, which may not provide a basis to predict the degree of comprehension ultimately achieved, as initial disruptions might improve comprehension by prompting increased focus, attention, and effort by the reader. Further, for some of the findings cited above (the delayed-keyword and summarization results) the interventions occurred well after any situation-model construction was complete. Thus, on the whole, these findings support the more general situation-model approach, which posits that accurate metacomprehension depends on cues produced by accessing and utilizing one's situation model.

Future work needs to continue to pursue the situation-model approach as well as the demands of monitoring approach to improving metacomprehension accuracy. It will be important to fully understand what contexts may allow access to correct cues for predicting comprehension as well as what contexts may support access to correct cues specifically in low-ability learners. More work is also needed to establish the relation between metacomprehension accuracy, self-regulation and learning. Thiede et al. (2003) showed that accuracy influenced the effectiveness of regulation and in turn overall reading comprehension. More recently Thomas and McDaniel (2007) have also provided some evidence that conditions that lead to better monitoring accuracy can also lead to better study choices and learning. However, these effects need to be replicated and extended to populations other than typical college students. It is not clear whether younger students or less-skilled readers can use their monitoring to guide regulation. Thus, future research should demonstrate how generally metacomprehension accuracy affects regulation of study and comprehension.

Author Notes

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305H030170 and R305B070018 to Keith Thiede, Thomas D. Griffin and Jennifer Wiley. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

- Anderson, M. C. M. & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica*, *128*, 110–118.
- Baker, J. M. & Dunlosky, J. (2006). Does momentary accessibility influence metacomprehension judgments? The influence of study-judgment lags on accessibility effects. *Psychonomic Bulletin and Review*, *13*, 60–65.
- Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*, 610–632.
- Butler, D. L. & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*, 245–281.
- Chi, M. T. H., DeLeeuw, N., Chiu, M., & Lavancher, C. (1994). Elicit self-explanation improves understanding. *Cognitive Science*, *18*, 439–477.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, *12*(5), 769–86.
- Dunlosky, J., Baker, J. M., & Rawson, K. A. (2006). Does aging influence people's metacomprehension? Effects of processing ease on judgments of text learning. *Psychology and Aging*, *21*, 390–400.
- Dunlosky, J. & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, *16*, 228–232.
- Dunlosky, J. & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*, 374–380.
- Dunlosky, J. & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory and Language*, *36*, 34–49.
- Dunlosky, J. & Rawson, K. A. (2005). Why does rereading improve metacomprehension accuracy? Evaluating the levels-of-disruption hypothesis for the rereading effect. *Discourse Processes*, *40*, 37–55.
- Dunlosky, J., Rawson, K., & Hacker, D. J. (2002). Metacomprehension of science text: Investigating the levels-of-disruption hypothesis. In J. Otero, J. A. Leon, & A. C. Graesser (Eds.), *The psychology of scientific text comprehension* (pp. 255–279). LEA, Mahwah, NJ.
- Dunlosky, J., Rawson, K. A., & McDonald (2002). Influence of practice tests on the accuracy of predicting memory performance for paired associates, sentences, and text material. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 68–92). New York, NY: Cambridge University Press.
- Dunlosky, J., Rawson, K. A., & Middleton, E. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language*, *52*, 551–565.
- Fischer, P. M. & Mandl, H. (1984). Learner, text variables, and control of text comprehension and recall. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 213–254). London: Lawrence Erlbaum Associates.
- Glenberg, A. M. & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 702–718.
- Glenberg, A. M. & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory and Cognition*, *15*, 84–93.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, *116*, 119–136.
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, *36*, 93–103.
- Hemmerich, J. & Wiley, J. (2002). Do argumentation tasks promote conceptual change about volcanoes. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

- Jee, B., Wiley, J., & Griffin, T. D. (2006) Expertise and the illusion of comprehension. *Proceedings of the Annual Conference of the Cognitive Science Society*.
- Kelemen, W. L. & Weaver, C. A. (1997). Enhanced memory at delays: Why do judgments of learning improve over time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1394–1409.
- King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology*, 93, 329–343.
- Kintsch, W. (1994). Learning from text. *American Psychologist*, 49, 294–303.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, 29, 133–159.
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100, 609–639.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370.
- Lefevre, N. & Lories, G. (2004) Text cohesion and metacomprehension: Immediate and delayed judgments. *Memory & Cognition*, 32, 1238–1254.
- Lin, L., Moore, D., & Zabrucky, K. M. (2001). An assessment of students' calibration of comprehension and calibration of performance using multiple measures. *Reading Psychology*, 22, 111–128.
- Lin, L. & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology*, 23, 345–391.
- Lin, L., Zabrucky, K. M., & Moore, D. (2002) Effects of text difficulty and adults' age on relative calibration of comprehension. *American Journal of Psychology*, 115, 187–198.
- Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10, 756–766.
- Maki, R. H. (1995). Accuracy of metacomprehension judgments for questions of varying importance levels. *American Journal of Psychology*, 108, 327–344.
- Maki, R. H. (1998a). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Hillsdale, NJ: LEA.
- Maki, R. H. (1998b). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition*, 26, 959–964.
- Maki, R. H. (1998c). Metacomprehension of text: Influence of absolute confidence level on bias and accuracy. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 223–248). San Diego, CA: Academic Press.
- Maki, R. H., Foley, J. M., Kajer, W. K. Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 609–616.
- Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin and Review*, 1, 126–129.
- Maki, R. H. & Serra, M. (1992a) The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 116–126.
- Maki, R. H. & Serra, M. (1992b). Role of practice tests in the accuracy of test predictions on text material. *Journal of Educational Psychology*, 84, 200–210.
- Maki, R. H., Shields, M., Wheeler, A. E. & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, 97, 723–731.
- McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, 38, 1–30.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1–43.
- Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, 131, 349–363.

- Nelson, T. O. & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed JOL effect." *Psychological Science*, 2, 267–270.
- Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation*, 26, pp. 125–141. New York: Academic Press, Inc.
- Nesbit, J. C. & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 76, 413–448.
- Olin, J. T. & Zelinski, E. M. (1997). Age differences in calibration of comprehension. *Educational Gerontology*, 23, 67–77.
- Pierce, B. H. & Smith, S. M. (2001). The postdiction superiority effect in metacomprehension of text. *Memory & Cognition*, 29, 62–67.
- Rawson, K. A. & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 69–80.
- Rawson, K. & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, 559–579.
- Rawson, K. A., Dunlosky, J., & McDonald, S. L. (2002). Influences of metamemory on performance predictions for text. *The Quarterly Journal of Experimental Psychology*, 55, 505–524.
- Rawson, K., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, 28, 1004–1010.
- Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1998). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science*, 6, 132–137.
- Shaughnessy, J. J. & Zechmeister, E. B. (1992). Memory monitoring accuracy as influenced by the distribution of retrieval practice. *Bulletin of the Psychonomic Society*, 30, 125–128.
- Stensvold, M. S. & Wilson, J. T. (1990). The interaction of verbal ability with concept mapping in learning from a chemistry laboratory activity. *Science Education*, 74, 473–480.
- Thiede, K. W. (1999). The importance of accurate monitoring and effective self-regulation during multitrial learning. *Psychonomic Bulletin & Review*, 6, 662–667.
- Thiede, K. W. & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, 28, 129–160.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95, 66–73.
- Thiede, K. W. & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1024–1037.
- Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 31, 1267–1280.
- Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (in press). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes*.
- Thomas, A. K. & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition*, 35, 668–678.
- Vesonder, G. T. & Voss, J. F. (1985). On the ability to predict one's own responses while learning. *Journal of Memory and Language*, 24, 363–376.
- Weaver, C. A. (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 214–222.
- Weaver, C. A., III & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition*, 23, 12–22.
- Weaver, C. A., Bryant, D. S., & Burns, K. D. (1995). Comprehension monitoring: Extensions of the Kintsch and van Dijk model. In C. A. Weaver, S. Mannes, & C. Fletcher (Eds.), *Discourse comprehension. Essays in honour of Walter Kintsch* (pp. 177–193). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Weaver, C. A. & Kelemen, W. L. (1997). Judgments of learning at delays: Shifts in response patterns or increased metamemory accuracy? *Psychological Science*, 8, 318–321.

- Weinstein, C. E. & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook on research in teaching* (3rd ed., pp. 315–327). New York: Macmillan.
- Wiley, J. (2001) Supporting understanding through task and browser design. *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Wiley, J., Griffin, T., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology*, 132, 408–428.
- Wiley, J., Griffin, T., & Thiede, K. W. (2008). To understand your understanding you must understand what understanding means. *Proceedings of Annual Conference of the Cognitive Science Society*.
- Wiley, J. & Voss, J. F. (1999) Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, 91, 301–311.
- Winne, P. H. & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts & P. Pintrich (Eds.), *Handbook of self-regulation*, pp. 531–566. New York: Academic Press, Inc.

7 The Importance of Knowing What You Know

A Knowledge Monitoring Framework for Studying Metacognition in Education

Sigmund Tobias

Institute for Urban and Minority Education, Teachers College,
Columbia University

Howard T. Everson

Fordham University and American Institutes for Research

Metacognition is one of the most widely studied constructs in contemporary psychological research in educational, instructional, and developmental psychology. Despite numerous attempts to define the construct of metacognition, agreement on how best to delimit it remains elusive (Hacker, Dunlosky, & Graesser, 1998; Tobias & Everson, 2000a). Nevertheless, the flood of research on metacognition over the past quarter century, both in the United States and elsewhere, suggests that murky definitions have not deterred investigators working in these fields. Earlier reviews of the literature on metacognition (see, for example, Metcalfe & Shimamura, 1994; Hacker et al., 1998) suggest that despite disagreement about its definition, metacognition is assessed typically from inferences based largely on observations of students' performance, or through interviews with students, or via self-report measures (Schraw & Impara, 2000). Thus, it appears that the method of observation defines the construct, that is, metacognition is the data from a metacognitive self-report inventory or from an observational protocol.

The research program dealing with metacognitive monitoring of students' prior knowledge described in this chapter was inspired by the seminal writings of both John Flavell (1979) and the late Ann Brown (1980; Palincsar, 2003), whose work focused on learning from instruction. Within this tradition, metacognition is defined as one's ability to monitor, evaluate, and make plans for one's learning (Tobias & Everson, 2000b). Also, following Pintrich, Wolters, and Baxter's work (2000), metacognition was further divided into three distinct components: knowledge about metacognition, the ability to monitor one's learning processes, and the meta-ability to control the learning processes (Tobias & Everson, 2002). The emerging consensus view, therefore, is that all three components of metacognition are necessary to regulate learning effectively and efficiently.

For more than a decade the metacognitive monitoring research program described in this chapter has aimed at extending and refining one major aspect of metacognition, assessing students' ability to monitor their prior knowledge, or what they have learned previously in a particular academic domain. More specifically, the focus of this research program has been to differentiate between what students know and do not know in various academic domains, from language arts (vocabulary and word knowledge) to mathematics and problem solving. The underlying rationale is straightforward: If students fail to differentiate

what they know or have learned previously from what they do not know or need to learn (or relearn), they are not expected to engage more advanced metacognitive strategies, such as evaluating their learning in an instructional setting, or employing more efficient learning and studying strategies.

The purpose of this chapter is to provide a brief overview of this research program, which to date includes 26 separate studies that have investigated students' ability to monitor their prior knowledge, relate the results to the body of research in this area, and to suggest further needed research. In general, the principal research design involves using a novel method, the knowledge monitoring assessment (KMA), to study systematically the empirical relationships between students' ability to monitor their knowledge states and their academic achievement, as well as the relationships with other important psychological constructs such as motivation, anxiety, and self-regulated learning. Throughout the course of this discussion, issues and concerns will be raised about the validity and reliability, generally, of metacognitive assessment instruments, including the KMA. In keeping with the general theme of this handbook, the chapter will conclude with a discussion of the importance of advancing knowledge about metacognition and how, through focused attention and training, research can help improve student learning.

A Knowledge Monitoring Framework

Like many complex psychological constructs, the successful (i.e., psychometrically sound) measurement of metacognition is challenging (Pintrich et al., 2000; Schraw & Impara, 2000; Schraw, this volume). By definition, metacognition is a higher-order, executive process that monitors and coordinates other cognitive processes engaged during learning, such as recall, rehearsal, or problem solving to name a few. Thus, measuring metacognition poses considerable challenges. As the literature suggests, metacognition has been assessed in two principal ways: first, by observing students' performances on cognitively complex tasks, and second through the use of self-report inventories. Both approaches raise important questions and issues about the validity of inferences based on these evidentiary approaches. Schraw and Impara (2000), for example, pointed out that developing effective methods to assess metacognition has been difficult. Such assessment methods have required detailed observation and recording of students' learning, rating the observations for metacognition, obtaining "think aloud" protocols of students' work, and rating transcriptions of the introspective reports. Royer, Cisero, and Carlo (1993) noted that these methods are labor-intensive and, therefore, limit the number of participants that can be used in such research. Available self-report scales of metacognition, on the other hand, though relatively inexpensive to develop and score, present issues of validity about the inferences drawn from simply asking students to self-assess subtle, and often relatively inaccessible, higher-order metacognitive processes. Is it safe to assume that students are aware of the metacognitive processes used during learning, or that those introspective judgments are accurate reflections of what goes on in students' heads as they learn new and unfamiliar subjects? Metacognition involves some very complex cognitive processes. Do students know, and can they recall and report accurately on these processes during learning, or recall them afterwards? More importantly, do students report candidly? While honesty is always an issue with self-reports, special concerns apply to metacognitive processes. Truthful and direct responses are often a source of discomfort to students and other fledgling learners. Students may be reluctant to admit that they often use superficial or ineffective time-saving strategies to manage their academic learning. As indicated elsewhere (Tobias, 2006), there is evidence to suggest that these problems limit the validity of self-report instruments.

Knowledge Monitoring Assessment

The metacognitive ability to monitor accurately one's prior knowledge is central to learning effectively and efficiently from instruction (Tobias & Everson, 2000a, 2000b, 2002), in or out of school, or in the workplace (Tobias & Fletcher, 2000). Learners who distinguish correctly between what they have learned, and what they have yet to learn, have an advantage during instruction because they can omit or skim the more familiar material, and concentrate more fully on the less familiar content they have yet to master. In contrast, learners with underdeveloped metacognition, i.e., those with less accurate knowledge monitoring skills, often spend too much time reviewing familiar material at the expense of mastering the unfamiliar or new material and, as a consequence, often fall behind in the instructional sequence.

The theoretical perspective of the role of metacognition in most educational contexts that drives the knowledge monitoring research program is illustrated in Figure 7.1. As the figure indicates, the ability to monitor and accurately assess one's knowledge state is a fundamental, foundational component of metacognition. It is much less likely that students could use more sophisticated, higher-order metacognitive activities such as planning, selecting appropriate strategies, or evaluating learning accurately, if they fail to distinguish the material they know and (or) have mastered from the content they do not know. Research reported by Tobias and Everson (2000a, 2000b, 2002), and discussed later in this chapter, has provided support for this assumption.

The KMA, therefore, was developed to allow for more precise calibrations of students' ability to monitor their prior learning, with an emphasis on the accuracy with which students distinguish between what they know or problems they could solve from what they do not know or problems they cannot solve (Tobias, Hartman, Everson, & Gourgey, 1991; Tobias & Everson, 2000a, 2000b, 2002). Within this theoretical and measurement framework, students are presented with challenging academic tasks, e.g., word knowledge questions, mathematical equations, mathematical word problems, scientific terms, verbal analogies, etc., and are required to estimate (by indicating "yes" or "no") whether they know the material presented. When given mathematical word problems, for example, students are asked whether they believe they can solve each of the problems, if given

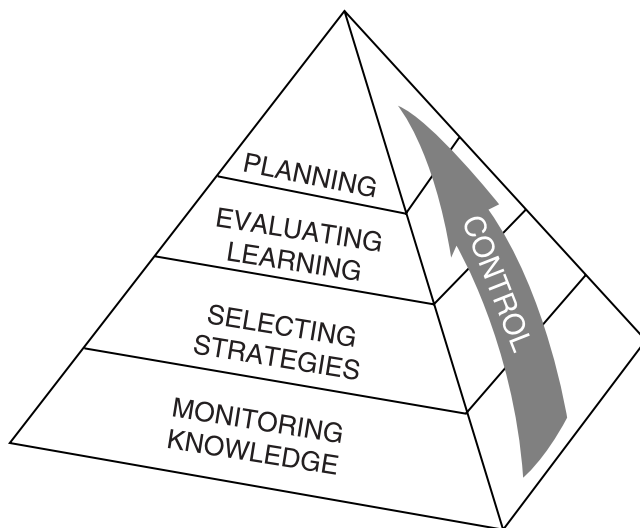


Figure 7.1 Hierarchy of metacognitive processes.

sufficient time. After making these estimates, students are then given multiple-choice test items on the same content. In this way, comparisons between the accuracy of students' metacognitive estimations and their subsequent test-based performance are used to support inferences about the accuracy of knowledge monitoring. This measurement approach is akin to methods used in research on metamemory (Hart, 1965; Koriat, 1994; Nelson, 1984; Nelson & Narens, 1994; Schraw, 1995), reading comprehension (Glenberg, Sanocki, Epstein, & Morris, 1987), and signal detection (Swets, 1986).

Like a number of other measures of metamemory, the KMA incorporates students' self-reports of their knowledge and (or) their ability to solve problems. However, unlike many other self-report measures, the KMA does not ask students to report on the cognitive processes they have used while performing cognitively complex tasks. It is expected that the KMA estimates are more accessible to students than are introspective reports of cognitive processes and, therefore, the KMA results are presumed to be less influenced by construct-irrelevant variance associated with recall and reporting. Gerrity and Tobias (1996) found that the KMA differentiated successfully between high-school dropouts and continuing students while test anxiety scales failed to distinguish between the groups, a finding attributed to the influences of social desirability factors. Moreover, Pintrich et al.'s (2000) review of research on existing metacognitive assessment instruments found that the KMA had the highest relationship with external criteria, such as reading ability, grade point average, or mathematics test scores, compared to other metacognitive assessment instruments.

Accurate Knowledge Monitoring

Obviously the accuracy of students' knowledge monitoring is key to the validity of the KMA. On the KMA, students' responses take the form of a 2x2 matrix with binary knowledge estimates (yes or no) as the columns and task or test item scores (also binary) represented by the rows (see Figure 7.2 for an example of the student response matrix). As Figure 7.2 indicates, four scores for each item/task presented to the student are produced indicating the number of items/tasks estimated as known and subsequently scored as correct on the test (+,+), estimated as unknown yet scored as correct (-,+), estimated as known and scored as incorrect (+,-), and estimated as unknown and scored as incorrect (-,-). These are then summed across all KMA items presented. Scores reflecting relative knowledge monitoring accuracy, then, are an additive function of the number of ++ scores and - scores.

A number of investigators, including Nelson (1984, 1996), and Nietfeld, Enders, and

		Knowledge Estimate	
		YES (+)	NO (+)
Test Item Score	CORRECT (+)	+ +	+ -
	INCORRECT (-)	- +	- -

Figure 7.2 Item estimate and scoring matrix.

Schraw (2006), have suggested that the optimal analysis of the discrepancies between estimated and demonstrated knowledge requires a probabilistic conceptualization of knowledge monitoring ability. They encourage the use of either the Gamma (γ) coefficient, a measure of association (Goodman & Kruskal, 1954), or the Hamman (η) coefficient, a measure of agreement accuracy (Romesburg, 1984). As noted earlier, these and similar methods have been used in metamemory research on the feeling of knowing and the accuracy of feeling-of-knowing judgments (Hart, 1965; Nelson, 1984; Nelson & Narens, 1994). Although there has been some debate about which of the two indices is more suitable (Nelson, 1996; Wright, 1996), Nietfeld et al. (2006) have argued that γ is less appropriate when the absolute accuracy of agreement is central, and when training to improve metacognitive accuracy is involved. Research by Schraw (1995), Schraw and Impara (2000) and Nietfeld et al. (2006) has demonstrated, and the work using the KMA reinforces this point, that γ may actually lead to distortions in the response data because it returns a value of 1 when there are empty cells in the 2x2 response matrix and, consequently, may result in somewhat misleading (biased) estimates of knowledge monitoring accuracy. In contrast, η is less susceptible because empty cells do not affect the value returned. For these reasons, and others described elsewhere (Tobias & Everson, 2002), the η coefficient was used as an estimate of knowledge monitoring accuracy in the last 14 knowledge monitoring studies summarized below.

Help Seeking

Seeking help or using other forms of instructional support when a learning impasse occurs are also useful behavioral indices of metacognition because they suggest students' recognition that their learning is in need of repair. A number of studies investigated the efficacy of help seeking in learning situations (Karabenick, 1998; Nelson-LeGall, 1981) and report that it is a useful strategy. There are some contradictory findings about the use of help. Alevén, Stahl, Schworm, Fischer, and Wallace (2003) reviewed the use of help in interactive learning environments to give guidance about navigating through a sequence, or providing additional explanations on the use of help in such environments. They found that "an increasing number of studies provide evidence that learners often do not use help functions very effectively or even ignore them totally" (Alevén et al., 2003, p. 278).

Recent versions of the KMA incorporate help by providing students with assistance when requested. In KMAs designed to measure vocabulary knowledge monitoring accuracy, for example, the help function allowed students to access dictionary definitions of the word and offered examples of its use in a sentence. Similarly, in mathematical knowledge monitoring studies the KMAs included a feature enabling students to see the correct answer and the steps used to derive the mathematical solution. For example, Njoku (2007) used a KMA with a help function to assess vocabulary word knowledge and found that the help feature was used an average of 56% of the time when answering test items. These data suggest that when help is available to address the challenges of specific test items, rather than more global or general forms of instructional support, students seek help more frequently than reported by Alevén et al. (2003).

As indicated elsewhere (Tobias, 2006), much of the research on help seeking has been conducted in classroom environments and data on the frequency of help seeking has been gathered through self-reports. This design confounds two separate components of help seeking: (a) recognition that an error was made or that a learning impasse has occurred for which assistance is needed, and (b) the students' desire or motivation to seek help to solve the problem. Thus, one advantage of using the KMA to study help seeking is that the two components can be disentangled. If students estimate correctly that they do not know how

to solve a problem (and ultimately fail to solve it), assistance is clearly needed. On the other hand, if students seek help after estimating that they know the material (or know how to solve the problem) presented, then the assistance provided is unnecessary and redundant. This additional information can be used to confirm the accuracy of students' metacognitive judgments, and may provide further insights into students' help seeking more generally.

Help in the form of guidance to students is relevant not only to the assessment of metacognition but it also plays a role in the controversy about the effectiveness of instruction, particularly when situated in the debates about the effectiveness of constructivist instruction (Kirschner, Sweller, & Clark, 2006, Tobias & Duffy, in press). Wise and O'Neill (in press) suggest that the term guidance is too broad and has included explanations, general help, modeling, scaffolding, procedural direction, feedback, and other forms of instructional support. Clark (in press) also indicates that the operational and lexical definitions of guidance vary widely and that many forms of assistance to students are referred to as guidance. Moreover, Wittwer and Renkl (2008) suggest that "the vast majority of empirical research has demonstrated that the provision of instructional explanations does not necessarily foster learning outcomes" (p. 52). To help resolve the debate, Tobias (in press) has suggested that different types of "guidance" are best conceptualized as different forms of instructional support, i.e., forms of assistance provided to students, and that a hierarchy of such support is urgently needed. Such a hierarchy may stimulate further research to clarify the contradictory findings regarding the use of help, such as those regarding help in KMA studies, explanations in most instructional settings (Wittwer & Renkl, 2008), and use of help in interactive learning environments (Aleven et al., 2003). A hierarchy would help to clarify the conditions for optimizing instructional support, while also informing research about whether these conditions interact with learner characteristics.

Overview of KMA Research

The results of the first 12 studies using the KMA have been described in detail elsewhere (Tobias & Everson, 2000a, 2000b), and will be discussed only briefly here. The general theme emerging from the results was that there were substantial relationships between the KMA and standardized achievement test results across a number of academic domains. For example, correlations with performance on reading comprehension tests and mathematics assessments were strong ($r = 0.67$ and 0.76 , respectively), indicating that the more metacognitively able students, i.e., those with more accurate knowledge monitoring abilities, had higher achievement. Significant relationships were also found between the KMA and grades in college and vocational school settings. Although grades are somewhat unreliable indicators of students' academic achievement (Werts, Linn, & Joreskog, 1978; Willingham, Lewis, Morgan, & Ramist, 1990), relationships were found, nevertheless, with grades obtained concurrently by students, as well as prospectively when the KMA was administered prior to starting college and correlated with first-year grades. Further support for the KMA came from finding a negative correlation between KMA scores and the need for feedback in elementary school children ($r = -0.79$) suggesting that students who could accurately differentiate between what they did and did not know required less external feedback about their work. Similarly, negative relationships between knowledge monitoring ability and mathematics anxiety confirmed expectations that anxious students demonstrate less accurate, less well developed metacognitive abilities.

Earlier studies also indicated that the KMA was useful for differentiating between students who had dropped out of high school (Gerrity & Tobias, 1996), and in identifying

students diagnosed as being learning disabled. Commenting on this initial program of research, the late Paul Pintrich and his colleagues (Pintrich et al., 2000) noted that these studies provided the strongest evidence of the relationships among metacognition and relevant external criteria reported thus far in the literature.

Knowledge Monitoring in Different Instructional Setting

The second phase of the metacognitive knowledge monitoring research program examined the generalizability of knowledge monitoring. Eleven studies investigated the relationship between metacognitive knowledge monitoring and academic achievement (including examination of the relationships with motivation, help seeking, and anxiety) across a number of academic domains and a variety of instructional settings. Detailed descriptions of these studies and a more exhaustive discussion of the results can be found in Tobias and Everson (2002). Only a brief summary of this work is provided below.

Knowledge Monitoring and Reading

Two studies investigated the relationship between knowledge monitoring and reading among elementary school students. In the first study, groups of good and poor readers in the fifth and sixth grades were asked to estimate their vocabulary knowledge and complete a multiple choice test that included the same words (the typical KMA approach). As expected, students in the higher reading groups were found to be more accurate knowledge monitors than the less able readers. The correlation between students' scores on a standardized reading test and a vocabulary-based KMA was high ($r = 0.62, p < 0.001$), replicating the results of the first study. This finding also converged with results from earlier studies using college students.

In the second study, which investigated reading in a group of fourth-grade students, participants had the opportunity to get help, consisting of a correct word definition and use of the word in a sentence for 50% of the vocabulary test items. In this instance, the KMA was administered individually using index cards, and the help feature appeared on the flip side of each card. It was reasoned that students who accurately distinguished what they knew and did not know were likely to seek help only when it was needed most, i.e., on words estimated as unknown and scored incorrect when tested. Results indicated that, as expected, more accurate knowledge monitors sought help more strategically, i.e., where it was needed most. However, they also sought more help than needed on items estimated as known and passed on test. These results may have been attributable to the procedures, which required participants to ask for help on half the items which was more than they wanted or needed. Succeeding research examined this possibility.

Knowledge Monitoring and Mathematics

A KMA consisting of 26 mathematics problems (one-half word problems and one-half computational problems) was administered to tenth-grade students. Using the typical KMA procedures, students estimated their ability to solve the problems, completed a multiple choice test on the same problem set, and then after each test item was answered they had the option to get help (i.e., they could view the correct answer as well as the steps to derive it). As in the preceding study, the differences in help seeking patterns suggested that students who were classified as accurate knowledge monitors sought help more often on those mathematics items they estimated as unknown and which they answered incorrectly (those with $-,-$ scores), while their less accurate peers sought help more often for

problems estimated as known and answered incorrectly (+,– scores). These help seeking differences were related, in part, to the fact that those with higher KMA scores estimated fewer items as being known and failed on test.

Knowledge Monitoring and General Academic Ability

Virtually all of the knowledge monitoring studies conducted that did not assess mathematics used vocabulary or verbal analogy items from prior SAT tests. Such items have always correlated highly with general intelligence; therefore significant relationships with intellectual ability were expected (Lemann, 1999; Lohman, 1997). A question examined in the earlier phase of KMA research was whether the accuracy of metacognitive monitoring contributed as much explanatory variance than word knowledge alone. In nearly three-fourths of the studies using verbal ability measures, knowledge monitoring accuracy contributed more variance “ranging in effect size or R^2 from 1% to 58% with a median of 4%” (Tobias & Everson, 2000a, p. 210) when compared to word knowledge alone. These results indicated that while word knowledge and other indicators of intellectual ability contributed variance to the results, accuracy of monitoring accounted for a significant amount of variance beyond these traditional indices of intellectual ability. More recent studies on KMA using the Hamann coefficient (η), where the metric is based on the proportional accuracy of estimating both known and unknown stimuli, rendered such comparisons unnecessary.

The relationship between knowledge monitoring and scholastic ability and aptitude were investigated in a number of studies. Sternberg’s (1998) triarchic intelligence tests were given to high-school students, who also took a KMA composed of words from tenth-grade curriculum materials. As expected, significant correlations were found between monitoring accuracy and analytic verbal intelligence ($r = 0.40$, $p < 0.01$). Unexpectedly, KMA scores also correlated with the practical, quantitative subtest ($r = 0.29$, $p < 0.05$). In view of the prior results relating knowledge monitoring to school learning, positive relationships with the analytic subscale were expected, though the second relationship could not be easily interpreted. In view of the fact that eight correlations were computed for this analysis, perhaps the latter correlation was obtained by chance.

The relationships between knowledge monitoring and scholastic aptitude were investigated by giving a KMA composed of verbal analogies, taken from prior versions of the SAT verbal reasoning test, to students attending a voluntary summer program for academically talented students ($N = 462$). Participants received another KMA composed of math items, also taken from prior versions of the SAT mathematical reasoning test and from the SAT II Mathematics Subject Test. As expected, SAT mathematics scores correlated ($r = 0.50$, $p < 0.001$) with a math KMA, while the SAT verbal scores correlated less strongly ($r = 0.29$, $p < 0.001$) with the verbal KMA scores.

In yet another study, 120 students in a pre-college course designed to help them succeed in their freshman year of college were given a vocabulary-based KMA, and one in mathematics also taken from prior SATs. Surprisingly, no significant relationships were found between students’ monitoring accuracy and either their grade point average (GPA) in the course, or with their SAT scores. Possibly instructors in this pre-college course may have rewarded students to encourage effort, despite limited attainment, accounting for the negative results.

In a military setting, a study relating the KMA to selected subtests of the Armed Forces Qualification Test (ASVAB) investigated the pattern of correlations in a sample of 71 U.S. Navy trainees. As expected, correlations between the KMA and scores on the ASVAB subtests measuring verbal ability were significant. These findings, along with the

preceding results, confirmed the expectation of moderate relationships between the KMA and various measures of scholastic aptitude.

Domain Specificity

The question of whether knowledge monitoring is domain specific or a more general ability was investigated in several studies. In the study with Navy personnel, KMA scores were generated from both general word knowledge and technical (oceanographic) word knowledge. The Navy personnel were studying content related to oceanography, though none of the words used in the KMA duplicated what was taught in the training course. Participants then read a three-page text passage in which all the oceanography words were defined either explicitly or implicitly. The two KMAs were then re-administered in order to obtain an index of trainees' ability to update their subject-specific knowledge and monitoring ability. The KMA data were then correlated with the trainees' post-test scores in the course they were taking. The general vocabulary KMA was significantly related ($r = 0.31$, $p < 0.01$) to the course post-test scores, as was the oceanography-based KMA—but only on its second administration, after the text passage was read.

In another study two KMAs, one with standard vocabulary items and a second with words used in a course dealing with learning theory, were given to 37 college students. Participants then completed an 11-item pretest based on a text dealing with learning theory (presented in hypertext format) which they read as a course requirement. After finishing the text a post-test, including pretest items, was given to assess their understanding of the text passage. Multiple regression analysis indicated that only the standard, non-technical KMA scores were related to post-test scores ($r = 0.72$, $p < 0.001$). The KMA employing words based on learning theory had no statistical relationship with the post-test scores.

In both of these studies correlations between general and domain-specific KMAs were low to moderate, but the KMA vocabulary items that were not domain specific appeared to be more strongly related to learning outcomes. Since learning in the courses students were taking, and in all school learning for that matter, relies at least in part on general knowledge, a relationship with general knowledge or intellectual ability (“g”) is obviously expected. Apparently words from the domain that is the subject of instruction, but not used in the courses, do not predict outcomes as well as more general words. A definitive study of the generality-specificity issue may require use of an entirely unfamiliar domain, and administration of two KMAs, one general and one from that particular domain, and then a comparison of learning in the novel and familiar domains.

Self-Reported Metacognition

The relationship between the KMA and self-reports of metacognition was examined in several studies. An initial study found low, non-significant relationships between students' college GPA and all the scales of the *Motivated Strategies Learning Questionnaire* (MSLQ) (Pintrich, Smith, Garcia, & McKeachie, 1991). In another, the *Learning and Study Skills Inventory* (LASSI) (Weinstein, Palmer, & Schulte, 1987) was also administered. Although neither the MSLQ nor the LASSI scores correlated significantly with college GPA, the correlation between grades and the KMA scores was significant. Furthermore, the correlations between the various self-report measures were relatively high, while the relationships between the KMA and the self-report questionnaires were not statistically significant, indicating that somewhat different characteristics were being measured by the various assessment instruments.

In another study the KMA had higher relationships with SAT scores than either the LASSI or the *Metacognitive Awareness Inventory* (Schraw & Dennison, 1994), a self-report scale. Again, substantial relationships between self-report measures of metacognition were found, but non-significant relationships among them and the KMA scores were also reported. These results indicate that the KMA and the scores derived from metacognitive self-report scales are quite low, and often do not reach statistical significance. These findings are not surprising, because the KMA was typically correlated with achievement and aptitude measures, while scores from self-report scales were generally unrelated to either, suggesting that the self-report scores are either assessing different constructs or have weaker relationships with external criteria.

In work done in Belgium, Van Damme and Griet (2001) also found generally low, though often statistically significant correlations between a vocabulary-based KMA and self-report measures of metacognition. They also asked teachers to rate students' metacognition and found high correlations between the KMA and teachers' judgments of "the extent to which the student has an accurate view on his own capacities and with the effectiveness of his study strategies" and concluded that the results "are without doubt strong indications for the validity of the KMA for the functioning of a student at school" (Van Damme & Griet, 2001, p. 3).

Motivation and Metacognition

The critical role played by motivation in school learning was succinctly summarized by Maehr and Meyer (1997) in their endorsement of former U.S. Secretary of Education Terrel Bell's statement that "There are three things to remember about education. The first is motivation. The second one is motivation. The third one is motivation" (Maehr & Meyer, 1997, p. 372). It is generally accepted that motivation initiates, directs, and maintains the activities controlling learning. However, motivation can affect learning only by engaging the metacognitive processes controlling it. Contemporary motivational research typically utilizes quasi-experimental designs in which motivation is assessed by self-report questionnaires and predictions are made about their reactions to a variety of situations. Motives may be seen as dispositions to achieve, or avoid, specific outcomes, desires, or wishes that give rise to the adoption of various goals (Pintrich & Maehr, 2001; McLelland, Koestner, & Weinberger, 1989). Research by Elliott and Church (1997) has shown that once goal orientations are taken into account, motivation has no residual effect. Motives may "arouse or activate certain types of goals, but the goals then serve to guide and direct behavior and achievement" (Pintrich & Maehr, 2001, p. 16).

Several KMA studies examined the effect of both metacognition and motivation, assessed by measures of motivational goal orientation. This research design employed a 2x2 conception of goals: mastery and performance, with each having both approach and avoidance valences. Mastery approach goals describe students' desires to learn and understand, while the avoidance element portrays their hope to evade misunderstanding or completing a task incorrectly. Performance approach goals, in contrast, describe the desire to outdo others, whereas the avoidance valence focuses on evading any suggestion of being less able than peers. It has been shown (Pintrich, 2000; Meece, Blumenfeld, & Hoyle, 1988) that more than one goal may be pursued simultaneously.

Tobias, Galvin, and Michna (2002) used a questionnaire developed by Pintrich and Maehr (2001) which yielded both mastery and performance approach and avoidance scores. Participants also completed a KMA consisting of 42 words administered by computer. On the first 21 words, students estimated their word knowledge of each word, answered a multiple choice item on that word, and then had the option to request help on

the item. The help feature consisted of receiving a definition of the word as well as an example of the word used in a sentence. On the second set of 21 words, the procedures were identical, except that feedback was given regarding whether or not the answer on the test items was correct; help could be requested on any word after feedback was provided. The results indicated that motivational goal orientation scores had no effect either on monitoring accuracy or on help seeking.

In a similar study, Stavrianopoulos (2004) investigated the effects of motivational goal orientation, induced by instructions and assessed by questionnaire, on metacognitive knowledge monitoring and help seeking. In this study Elliott and Church's (1997) scale was used and yielded scores for mastery, performance approach and avoidance. Mastery, performance approach and performance avoidance instructions, whose content validity was confirmed by judgments of leading researchers in this area, were used to induce students to complete the KMA. A control group also completed the KMA, but received motivationally neutral instructions. During debriefing, interview results indicated that students generally worked in accord with the instructions they received. Stavrianopoulos used 40 authentic vocabulary words in her KMA plus five nonsense words in an effort to determine whether more accurate knowledge monitors readily detected the nonsensical words. Participants ($N = 137$) were limited to requesting help on only ten words. It was thought that limiting the amount of help would make the experimental situation more comparable to school learning where students often do not have the time to seek help on every aspect of their school work. Hence, they need to be strategic in selecting help with the vocabulary words that gave them the greatest difficulty. In this research design, "help" consisted of providing a definition of the word and an example using the word in a sentence.

The results reported by Stavrianopoulos indicated that neither the motivational instructions, nor questionnaire scores, nor their interaction, affected students' knowledge monitoring accuracy. As expected, more accurate monitors sought help on more nonsense words than their less accurate counterparts, because, obviously, they were unfamiliar with these nonsensical words. Also as expected, more accurate knowledge monitors sought more help on words they had estimated as not knowing and, indeed, failed on the test. Thus, while supporting predictions dealing with metacognitive knowledge monitoring, Stavrianopoulos' results failed to confirm expectations dealing with motivational goal orientation.

Cross-Cultural Studies of Motivation and Metacognition

Tobias, Njoku, and Everson (2001) compared high school students from Nigeria ($N = 77$ males) to U.S. students from a middle-class community ($N = 70$, 39 were female); all students attended Catholic Parochial schools. In order to equalize the number of males in the Nigerian and American samples a second group of male participants ($N = 57$) was recruited from another U.S. high school, attended only by male inner city students. A KMA composed of 24 mathematics word problems, selected by experts as being equally appropriate for Nigerian and American students, was administered and participants had the option of receiving help consisting of the correct answer to the problem and the steps to reach the answer.

The results indicated that the middle-class American students were more accurate monitors than urban students who were classified as coming from families of lower socio-economic status. The demographic differences among the students permitted analyses of performance for the four groups (Nigerian males, and the three American groups which included middle class males, lower socio-economic status males, and middle class females).

Not surprisingly, substantial differences in monitoring accuracy were found among the four groups. Middle class participants were more accurate monitors than the Nigerian students and the U.S. students from the lowest socio-economic status group. Moreover, the middle class American males were marginally more accurate than the Nigerian students. No help seeking differences were found between the good and poor knowledge monitoring groups, though differences between the Nigerian and American students did occur with respect to seeking help on two types of KMA items: those estimated as unknown and failed on test (-,-), and those estimated as unknown and passed on test (-,+), with American students, generally, seeking more help. Though the sample sizes were small, further analyses suggested that differences in socio-economic status may account for more variance in performance than cross-cultural differences.

In a follow-up study Njoku (2007) used the KMA math problems from the earlier study and found that American students ($N = 164$) were more accurate knowledge monitors than their Nigerian counterparts ($N = 166$). Moreover, Njoku also reported differences in help seeking behavior, with Nigerian students seeking more help on two types of problems: those estimated to be unknown and failed on test (-,-) and those estimated as unknown and passed on test (-,+). Regression analyses indicated that, in general, more accurate knowledge monitors, irrespective of cultural background, were more strategic in seeking help. That is, more help was sought on items estimated as unknown and failed on test (-,-), and less help was sought for items estimated as known and failed (+,-). Unexpectedly, more accurate knowledge monitors also sought help more often on items estimated as known and passed on the test (+,+).

Njoku also administered Pintrich and Maehr's (2001) motivational goal orientation questionnaire. Students scoring high on Mastery Approach tended to be more accurate knowledge monitors; and that relationship was substantially higher for American students ($r = 0.46, p < 0.01$) than for those from Nigeria ($r = 0.16, p < 0.05$). In the Nigerian sample, Mastery Approach scores were found to be related to the amount of help sought on items estimated as unknown and failed on test (-,-), whereas in the American sample Mastery Approach scores were found to be associated with help sought for items estimated as known and passed on the test (+,+).

In summary, with the exception of the Njoku (2007) study, the research found few of the relationships between mastery motivational orientation and metacognition suggested by motivational theory. Unfortunately, space constraints make a more complete summary of this voluminous research impossible. However, a more detailed review of metacognitive and motivational goal orientation research is available elsewhere (Tobias, 2006).

Discussion

The findings summarized in this chapter suggest considerable support for the metacognitive knowledge monitoring construct and the use of the KMA as a method of assessing it. KMAs were developed using a reasonable range of content from vocabulary, mathematics (both word and computational problems), technical knowledge, to verbal analogies. Participants in the KMA research program have included students at all levels (elementary, junior high school, high school, and college) and of different characteristics, i.e., mainstream students, those diagnosed as learning disabled or having attention deficit disorders, high school dropouts, vocational high school students, college students, and military personnel. The variation in content and the diversity of participants strengthen the validity of the knowledge monitoring construct and the KMA methodology.

The low and sometimes non-significant relationships between the KMA and other metacognitive assessments suggest that somewhat different things are measured by the two

types of assessments. Furthermore, since the KMA was found to have higher relationships with external criteria than other types of assessments (Pintrich et al., 2000), the metacognitive abilities tapped by the KMA appear to be more closely related to school learning than those measured by other assessment tools, a conclusion also reached by Van Damme and Griet (2001). Together, these findings suggest that the KMA is a useful measurement tool for research into the complex aspects of metacognition.

Motivation

In studies investigating metacognitive knowledge monitoring and motivation, the findings dealing with the effects of motivation were largely disappointing, although the results for metacognition generally confirmed expectations with respect to more frequent detection of non-sense syllables, and more strategic use of help. Perhaps the differences in results for the two constructs are due, in part, to the way they were assessed; concerns about the assessment of motivation are described more fully elsewhere (Tobias, 2006).

The KMA is partially a performance-based measure. That is, participants have to demonstrate their knowledge as part of the assessment process. Relationships between metacognition and motivation may be different if the assessment of motivation relied more on performance measures as opposed to self-reports. A performance measure of motivation, such as persistence on a complex task, may be more revealing of individual differences in motivation. Another method for assessing motivation may be to give participants the opportunity to select the task they want to work on from a variety of alternative challenges. For example, in research on students' ratings of instructors Tobias and Hanlon (1975) reasoned that a useful measure of students' feelings about an instructor would be whether they intended to select the same instructor for another course in the same domain. A pseudo pre-registration procedure was implemented so that students could express their intention to register for the next course.

A meta-analysis of the behavioral intention literature (Webb & Sheeran, 2006) showed that a medium-to-large change in intention leads to a small-to-medium change in behavior. One suggestion for further research on motivation is to implement procedures in which students could declare specific intentions, perhaps by choosing among a number of alternatives, to engage in an action of some consequence to them. For example, participants might be offered alternatives for selecting among a series of classes in some of which they are likely to get good grades, while in others grades are de-emphasized but they would learn a good deal about the domain. Because intentions have been shown to be highly related to actual behavior, such choices might be more revealing about mastery or performance orientations than answering a multiple choice questionnaire.

Motivational Goal Orientation

In general, as indicated earlier, the research reported above confirmed expectations for metacognitive knowledge monitoring but often failed to support hypotheses from motivational goal orientation theory. Brophy (2005) noted related concerns about motivational goal orientation and its assessment generally, though his critique focused largely on performance goals. While these results are at variance with findings dealing with goal orientation assessed by self-reports (Elliot & McGregor 2001; Pintrich & Maehr, 2001; Zuscho & Pintrich, 2000; Harackiewicz, Barron, & Elliott, 1998; Finney & Davis, 2003; Barron, Finney, Davis, & Owens, 2003), the findings are similar to results reported when students can freely express their feelings and thoughts (Light, 2001; Nathan, 2006).

Light (2001), for example, reported on interviews, lasting from one to three hours, of

more than 1600 undergraduates to discuss getting the most out of college; many were interviewed more than once. Even though the interviewees were from one university, Harvard University, the results and conclusions were presented at more than 90 campuses where students and faculty agreed that the findings on their campuses would be similar. "A large majority of students say they learn significantly more in courses that are highly structured, with relatively many quizzes and short assignments. . . . In contrast, students are frustrated and disappointed with classes that require only a final paper" (Light, 2001, p. 8). While it is possible that some students prefer such procedures to attain mastery, these descriptions sound like the types of courses preferred by performance oriented rather than mastery students. The high frequency of these preferences raises questions about how important mastery motivation is to students. Similar findings were reported by Nathan (2006).

The late Michael Pressley and his colleagues conducted a series of studies dealing with students' college coping strategies using ethnographic interviewing, coupled with grounded theory procedures, and validated the model using a range of confirmatory methods (e.g., Pressley, Van Etten, Yokoi, Freebern, & Van Meter, 1998; Van Etten, Freebern, & Pressley, 1997; Van Etten, Pressley, Freebern, & Echevarria, 1998). The major theme emerging from their data was that achieving good grades was an overarching student goal. "Students made it quite clear that all other goals were secondary" (Pressley et al., 1998, p. 353). Mastery motivation was rarely mentioned by students.

A further concern about results emerging from the use of motivational goal orientation questionnaires is that they often use relatively large samples. In KMA studies, the most positive goal orientation results emerged from Njoku's (2007) dissertation study in which 330 students participated. Similar findings have been reported in the motivational goal orientation literature. Pintrich and Maehr (2001) used 458 and 203 subjects, respectively, in their two studies, while Finney and Davis (2003) employed 1,200 students in their research.

The findings reported above raise major questions about students' self-reports of mastery orientations in academic contexts. One possibility is that mastery motivation is a construct accounting for limited variance and attains statistical significance mainly in studies using relatively large samples. Another possibility is that when students respond to questionnaire items dealing with mastery motivation, they do so in what they perceive to be socially desirable directions and only profess to have mastery motivational tendencies. When students are free to verbalize anything they wish, however, mastery motivation is barely mentioned. Perhaps both alternatives, i.e., that mastery accounts for very little variance and that students respond in socially desirable ways, are plausible. In any event, the findings raise questions about the degree to which students' self-reports reflect accurately what types of goals are important to them. Finally, the findings regarding motivational goal orientation confirm the importance of using other measures of constructs, such as behavioral intentions and performance-based indices, rather than relying exclusively on self-reports.

Self-Regulation and Metacognition

Self-regulation is clearly an important and actively investigated construct (Pintrich, 2000; Pintrich et al., 2000; Pintrich & Zusho, 2002; Zimmerman, 1995, 2000, this volume) that is composed of both metacognition and motivation. Hong, O'Neil, and Feldon's (2005) findings that both metacognitive and motivational higher-order factors emerged in their research confirms the joint importance of these two constructs in self-regulated learning. Bandura (2001) also described the importance of self-regulation in many activities from

managing one's health, dealing with stressors, recovering from substance abuse, to acting altruistically in turbulent times. In all of these settings, being able to monitor one's behavior, evaluate its effectiveness, plan, and select strategies to accomplish one's purposes are important. Similarly, a desire or a need to understand a situation and avoid errors and mistakes are also important in accomplishing one's goals.

From a research perspective, however, it may be more fruitful to investigate metacognition and motivation as separate and distinct constructs, even though they may jointly affect self-regulated learning. In the KMA research studying both metacognition and motivation, both constructs were assessed and, as previously noted, metacognition often had the expected effects on learning and achievement, but motivational goal orientation, assessed by widely used self-report measures, did *not*. It seems likely that had measures of self-regulated learning been used, often assessed with specially prepared scales such as Bandura's (1989) self-regulated learning subscale, the trait and state scales developed by O'Neil and his colleagues (Hong, O'Neil, & Feldon, 2005; O'Neil & Abedi, 1996), the structured interview method (Zimmerman & Martinez-Pons, 1986) or with the MSLQ (Pintrich et al., 1991) that combines both metacognitive and motivational elements, fewer effects would have been found since the motivational component often turned out to be insignificant. Thus, if these results are replicated, i.e., that metacognition accounts for more variance than motivation, future research on self-regulated learning may benefit from unpacking these constructs, and treating them as distinct rather than assessing them as a single, unidimensional construct.

In retrospect, the finding that metacognition accounts for more variance in learning than motivation is not surprising. In order to react to situations effectively, people first need to determine that some course of action (adjusting, correcting, or preventing) is necessary. Of course, such a determination stems from metacognitive, not motivational, processes. While motivation may have a small role in initiating such appraisals, it becomes important mainly *after* the recognition that an impasse has occurred. As Bandura (2001, p. 21) indicated, "personal well-being is best served by highly accurate self appraisal in managing a variety of difficult situations." Findings from research on the monitoring of prior knowledge suggest that the major variance in such self-appraisals is contributed by metacognition.

Motivation and metacognition may be particularly important in novel learning environments, such as web-based learning (Tobias, 2006). In such contexts, including Advanced Distributed Learning (Fletcher & Tobias, 2003) and anytime-anywhere learning (Fletcher, Tobias, & Wisner, 2007), students have less instructional support from colleagues or instructors, receive less direction or monitoring than in school settings, and complete their learning largely independently. Thus, it is reasonable to expect that metacognitive and motivational processes would be more important in such settings than in traditional school contexts. Future researchers should consider the use of behavioral intentions or performance-based measures of motivation, such as some of those described above, in such research.

Future Directions

In the overwhelming number of metacognitive monitoring of prior knowledge studies reported above, students invariably estimated that they knew and/or could solve a particular KMA item (from 47% to 69%). Thus, even though the KMA provides a measure of the accuracy of students' metacognitive estimates, actual content knowledge and metacognitive accuracy are combined in this measurement paradigm. Future research should use a larger proportion of more challenging stimuli, i.e., that would result in estimates of "do not know/cannot solve." Obviously, adding items consisting of nonsense words or

including insoluble mathematical problems would be one way of doing that. When confronted with such items, students, especially those who are quite knowledgeable, will have to make estimates that they do not know an item and have it scored incorrectly, thereby helping to disentangle the possible contributions of knowledge and estimation of that knowledge.

From a psychometric perspective, studies of large-scale assessments (Ackerman, Gierl, & Walker, 2003; Leighton, Gokiert, & Cui, 2007) have shown that many, if not most, are multi-dimensional. Based on this work, it seems useful to examine the dimensionality of the KMA. Like many psychological and educational measurements, it seems likely that the KMA is inherently multidimensional—one dimension probably captures students' confidence regarding their knowledge or problem-solving ability, another may be their ability to discriminate the known from the unknown, and another (which may also have multiple dimensions) may represent students' knowledge in a particular domain. Developments in psychometrics provide a number of empirical methods for investigating the dimensionality of assessment data (Ackerman et al., 2003; Leighton et al., 2007) that future researchers may wish to utilize.

Similarly, further research on the relationship between knowledge monitoring, motivation, and conation is needed to determine their impact on students' ability to act as strategic learners. It would be useful to identify students who are accurate knowledge monitors, but not highly motivated to succeed academically. As a result of their knowledge monitoring skills, such students may well have effective test taking strategies and therefore succeed on some evaluations, though they may not be able to control their studying effectively or do well on tasks requiring sustained effort. On the other hand, highly motivated students who are willing to invest considerable effort to attain academic goals are unlikely to be effective unless they also possess accurate knowledge monitoring skills. Knowledge monitoring in general, and the KMA in particular, are less likely to be affected by motivational and volitional variables than are more complex self-regulatory behaviors because making knowledge or problem solving estimates is relatively automatic and takes little effort. Therefore, motivation and conation should have less influence on such judgments. On the other hand, obtaining additional help and/or reviewing prior learning is more effortful and, therefore, may be more affected by motivational and volitional tendencies.

It may also be helpful to obtain "think aloud" protocols during the estimation phase of the KMA procedure, much the way they were used by Moos and Azevedo (2008), Azevedo, Moos, Greene, Winters and Cromley (2008), and by Desoete (2007). Such protocols may clarify students' thinking as they estimate their knowledge. It would be especially interesting to determine whether students have any doubts while making estimates that turn out to be incorrect, i.e., that they know an item and fail it on the test, or do not know and subsequently pass the item. Doubt about such items may indicate fragmentary knowledge that could be utilized to train students to improve the accuracy of their future knowledge estimates, as well as to guide students to seek help in order to succeed on those items in the future.

A hierarchical organization of metacognitive activities was assumed by the research described above. That is, it was assumed that more advanced metacognitive abilities such as evaluating learning, selecting appropriate learning strategies, or planning future learning activities could not occur without accurate knowledge monitoring. Support for that position may be inferred from the research results; nevertheless, the hypothesized hierarchy of metacognitive processes should be subjected to empirical test. Thus, it would be interesting to observe students in classes while they conduct self-evaluations of their work, select strategies, or plan for future learning to determine whether accurate knowledge monitors do, in fact, conduct these activities more efficiently than other students. The

significant relationships between teachers' judgments and students' KMAs reported by Van Damme and Griet (2001) suggest this may be a fruitful avenue for further research.

Finally, it has been assumed that the KMA is more resistant to the problems posed for self-report instruments in general, such as students' tendency to make socially desirable responses. That assumption, also supported by the research results, should be studied more thoroughly by instructing some students to respond to the KMA in ways that present them as diligent students, while others are instructed to appear casual or lackadaisical; of course, a control group should receive instructions telling them only how to complete the task. In such studies, participants would also be asked to respond to self-report measures to study the relative resistance of both types of assessment to student "faking."

Training Metacognitive Knowledge Monitoring

The findings of a number of studies indicating that accurate knowledge monitoring was related to achievement in different domains demonstrated the importance of accurate knowledge monitoring in a variety of school settings. A useful suggestion for further research is to determine if students' knowledge monitoring accuracy can be improved by training. It has been shown (Desoete, this volume) that metacognitive processes can be improved by training, suggesting that similar efforts to improve the monitoring accuracy of students' prior knowledge are also likely to be successful. Future research may then examine whether implementing knowledge monitoring training improves students' school learning. In addition to demonstrating the importance of knowledge monitoring, such research may lead to the development of effective interventions in a variety of areas.

References

- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement, Issues and Practice*, 22 (3) 37–51.
- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73, 277–320.
- Azevedo, R., Moos, D. C., Greene, J. G., Winters, F. I., & Cromley, J. C. (2008). Why is externally regulated learning more effective than self-regulated learning with hypermedia. *Educational Technology Research and Development*, 56, 45–72.
- Bandura, A. (1989). *The multidimensional self efficacy scales*. Unpublished test. Stanford CA: Stanford University.
- Bandura, A. (2001). The changing face of psychology at the dawning of a globalization era. *Canadian Psychology*, 42, 12–42.
- Barron, K. E., Finney, S. J., Davis, S. L., & Owens, K. M. (2003, April). *Achievement goal pursuit: Are different goals activated and more beneficial in different types of academic situations?* Paper delivered at the annual convention of the American Educational Research Association, Chicago IL.
- Brophy, J. (2005). Goal theorists should move on from performance goals. *Educational Psychologist*, 40, 167–176.
- Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. B. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453–481). Hillsdale, NJ: Erlbaum.
- Clark, R. E. (in press). How much and what type of guidance is optimal for learning from instruction? In S. Tobias & T. D. Duffy (Eds.), *Constructivist instruction: Success or failure?* NY: Routledge.
- Desoete, A. (2007). Evaluating and improving the mathematics teaching-learning process through metacognition. *Electronic Journal of Research in Educational Psychology*, 5, 705–730.

- Elliot, A. & Church, M. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 72, 218–232.
- Elliot, J. & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, 3, 501–519.
- Finney, S. J. & Davis, S. L. (2003, April). *Examining the invariance of the achievement goal questionnaire across genders*. Paper presented at the annual convention of the American Educational Research Association, Chicago IL.
- Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34, 906–911.
- Fletcher, J. D. & Tobias, S. (2003). *Implications of advanced distributed learning for education* (Urban Diversity Series). New York: ERIC Clearinghouse on Urban Education, Teachers College, Columbia University. Retrieved from http://www.iume.tc.columbia.edu/eric_archive/mono/UDS118.pdf.
- Fletcher, J. D., Tobias, S., Wisner, R. A. (2007). Learning anytime, anywhere: Advanced distributed learning and the changing face of education. *Educational Researcher*, 36 (2), 96–102.
- Gerrity, H. & Tobias, S. (1996, October). *Test anxiety and metacognitive knowledge monitoring among high school dropouts*. Paper presented at the annual convention of the Northeastern Educational Research Association, Ellenville NY.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, 116, 119–136.
- Goodman, L. & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association*, 49, 732–764.
- Hacker, D. J., Dunlosky, J., & Graesser, A. C. (1998). *Metacognition in educational theory and practice*. Mahwah NJ: Erlbaum.
- Harackiewicz, J. M., Barron, K. E., & Elliot, A. J. (1998). Rethinking achievement goals: When are they adaptive for college students and why? *Educational Psychologist*, 33, 1–21.
- Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56, 208–216.
- Hong, E., O’Neil, H. F., Jr., & Feldon, D. (2005). Gender effects on mathematics achievement: Mediating role of state and trait self-regulation. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics* (pp. 264–293). Cambridge, MA: Cambridge University Press.
- Karabenick, S. A. (1998). Help seeking as a strategic resource. In S. Karabenick (Ed.), *Strategic help seeking: Implications for learning and teaching* (pp. 1–12). Mahwah, NJ: Erlbaum.
- Kirschner, P. A., Sweller, J., & Clark, R. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential and inquiry-based teaching. *Educational Psychologist*, 41, 75–86.
- Koriat, A. (1994). Memory’s knowledge of its own knowledge: The accessibility account of the feeling of knowing. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 115–136). Cambridge, MA: MIT Press.
- Leighton, J. P., Gokiert, R. J., & Cui, Y. (2007). Exploratory and confirmatory methods to identify the cognitive dimensions in a large-scale science assessment. *International Journal of Testing*, 7(2), 141–189.
- Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. NY: Farrar, Straus and Giroux.
- Light, R. J. (2001). *Making the most of college: Students speak their minds*. Cambridge MA: Harvard University Press.
- Lohman, D. F. (1997). Lessons from the history of intelligence testing. *International Journal of Educational Research*, 27, 1–20.
- Maehr, M. L. & Meyer, H. A. (1997). Understanding motivation and schooling: Where we’ve been, where we are, and where we need to go. *Educational Psychology Review*, 9, 371–409.
- McLelland, D., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, 96, 690–702.
- Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students’ goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, 80, 514–523.

- Metcalfe, J. & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press.
- Moos, D. C. & Azevedo, R. (2008, April). *Metacognition and learning with hypermedia: To what extent do prior domain knowledge and self-efficacy matter?* Paper presented at the annual convention of the American Educational Research Association, New York.
- Nathan, R. (2006). *My freshman year: What a professor learned by becoming a student*. London: Penguin Books.
- Nelson, T. O. (1984). A comparison of current measures of accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109–133.
- Nelson, T. O. (1996). Gamma is a measure of accuracy of predicting performance on one item relative to another item, not the absolute performance of an individual item: Comments on Schraw (1995). *Applied Cognitive Psychology*, 10, 257–260.
- Nelson, T. O. & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–27). Cambridge, MA: MIT Press.
- Nelson-Le Gall, S. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review*, 1, 224–246.
- Nietfeld, J. L., Enders, C. K., & Schraw, G. (2006). A monte carlo comparison of measures of relative and absolute monitoring accuracy. *Educational and Psychological Measurement*, 66 (2), 258–271.
- Njoku, H. U. (2007). *The interrelations of goals, knowledge monitoring, strategic help seeking, and achievement among high school students of different cultures*. Unpublished doctoral dissertation, Fordham University, NY.
- O’Neil, H. F. & Abedi, J. (1996). Reliability and validity of a state metacognitive inventory: Potential for alternative assessment. *Journal of Educational Psychology*, 89, 234–245.
- Palincsar, A. S. (2003). Ann L. Brown: Advancing a theoretical model of learning and instruction. In B. J. Zimmerman & D. H. Schunk (Eds.), *Educational psychology: A century of contributions* (pp. 459–475). Mahwah, NJ: Erlbaum.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekarts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.
- Pintrich, P. R. & Maehr, M. L. (2001, April). *Motives, goals, and context: Sources for motivation, self-regulation, and achievement*. Paper presented at the convention of the American Educational Research Association, Seattle WA.
- Pintrich, P. R., Smith, D. A., Garcia, T., McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.
- Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–98). Lincoln, NE: Buros.
- Pintrich, P. R. & Zusho, A. (2002). The development of self regulation: The role of cognitive and motivational factors. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 249–284). San Diego, CA: Academic Press.
- Pressley, M., Van Etten, S., Yokoi, L., Freebern, G., Van Meter, P. (1998). The metacognition of college studentship: A grounded theory approach. In D. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 347–366). Mahwah, NJ: Erlbaum.
- Romesburg, H. C. (1984). *Cluster analysis for researchers*. London: Wadsworth.
- Royer, J. M., Cisero, C. A., & Carlo, M. (1993). Techniques and procedures for assessing cognitive skills. *Review of Educational Research*, 63, 201–243.
- Schraw, G. (1995). Measures of feeling-of-knowing accuracy: A new look at an old problem. *Applied Cognitive Psychology*, 9, 321–322.
- Schraw, G. & Denison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19, 460–475.
- Schraw, G. & Impara J. C. (2000). *Issues in the measurement of metacognition*. Lincoln NE: Buros.
- Stavrianopoulos, K. (2004). *Achievement goals as predictors of metacognitive knowledge monitoring*

- and academic help seeking behavior among college students. Unpublished doctoral dissertation, Fordham University NY.
- Sternberg, R. J. (1998). *The triarchic mind: A new theory of human intelligence*. NY: Viking Press.
- Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, 99, 100–117.
- Tobias, S. (2006). The importance of motivation, metacognition, and help seeking in web based learning. In H. F. O’Neil, Jr. & R. S. Perez (Eds.), *Web-based learning: theory, research and practice* (pp. 203–220). Mahwah NJ: Lawrence Erlbaum Associates.
- Tobias, S. (in press). An eclectic appraisal of the success or failure of constructivist instruction. In S. Tobias & T. D. Duffy (Eds.), *Constructivist instruction: Success or failure?* NY: Routledge.
- Tobias, S. & Duffy, T. D. (in press) *Constructivist instruction: Success or failure?* NY: Routledge.
- Tobias, S. & Everson, H. T. (2000a). Cognition and metacognition. *Educational Issues*, 6, 167–173.
- Tobias, S. & Everson, H. T. (2000b). Assessing metacognitive knowledge monitoring. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 147–222). Lincoln, NE: Buros.
- Tobias, S. & Everson, H. T. (2002). *Knowing what you know and what you don’t: Further research on metacognitive knowledge monitoring* (College Board Rep. No. 2002–03). New York: College Board.
- Tobias, S. & Fletcher, J. D. (Eds.). (2000). *Training and retraining: A handbook for business, industry, government, and the military*. New York: Macmillan Gale Group.
- Tobias, S., Galvin, K., & Michna, G. (2002, April). *Motivation, metacognition, and help seeking*. Paper presented at the annual convention of the American Educational Research Association convention, New Orleans, LA.
- Tobias, S. & Hanlon, R. (1975). Attitudes towards instructors, social desirability and behavioral intentions. *Journal of Educational Psychology*, 67, 405–408.
- Tobias, S., Hartman, H., Everson, H., & Gourgey, A. (1991, August). *The development of a group administered, objectively scored metacognitive evaluation procedure*. Paper presented at the annual convention of the American Psychological Association, San Francisco, CA.
- Tobias, S., Njoku, H., & Everson, H. (2001, January). *Cross cultural research on metacognitive knowledge monitoring and help seeking*. Paper presented at the annual meeting of the Winter Conference on Discourse, Text, and Cognition, Jackson Hole, WY.
- Van Damme, J. & Griet, T. (2001, April). *Relationships between metacognitive knowledge monitoring, self-report measures, and teacher judgments*. Paper delivered at the annual meeting of the American Educational Research Association, Seattle, WA.
- Van Etten, S., Freebern, G., & Pressley, M. (1997). College students’ beliefs about examination preparation. *Contemporary Educational Psychology*, 22, 192–212.
- Van Etten, S., Pressley, M., Freebern, G., & Echevarria, M. (1998). An interview study of college freshmen’s beliefs about their academic motivation. *European Journal of Psychology of Education*, 13, 105–130.
- Webb, T. L. & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, 132, 249–268.
- Weinstein, C. E., Palmer, D. R., & Schulte, A. C. (1987). *LASSI: Learning and study strategies inventory*. Clearwater, FL: H & H Publishing.
- Werts, C., Linn, R. L., & Joreskog, K. G. (1978). Reliability of college grades from longitudinal data. *Educational and Psychological Measurement*, 38, 89–95.
- Willingham, W. W., Lewis, C., Morgan, R., & Ramist, L. (1990). *Predicting college grades: An analysis of institutional trends over two decades*. NY: The College Board.
- Wise, A. F. & O’Neill, J. K. (in press). Making knowledge useful: The missing parts of the equation. In S. Tobias & T. D. Duffy (Eds.), *Constructivist instruction: Success or failure?* New York: Routledge.
- Wittwer, J. & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, 43, 49–64.

- Wright, D. B. (1996). Measuring feeling of knowing: Comment on Schraw (1995). *Applied Cognitive Psychology, 10*, 261–268.
- Zimmerman, B. J. (1995). Self-regulation: A social cognitive perspective. *Educational Psychologist, 30*, 217–221.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekarts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.
- Zimmerman, B. J. & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal, 23*, 614–628.
- Zuscho, A. & Pintrich, P. R. (2000, April). *Fear of not learning? The role of mastery avoidance goals in Asian American and European American college students*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Part IV

Writing

8 Metacognition and Children's Writing

Karen R. Harris, Steve Graham, Mary Brindle, and Karin Sandmel
Vanderbilt University

Writing today is not a frill for the few, but an essential skill for the many.

The National Commission on Writing

Writing is a critical component of humanity's story, so much so that historians distinguish between pre-history and history, with history defined by the invention of writing over 5,000 years ago. Today, approximately 85% of the world's population writes (Swedlow, 1999). Once used primarily for record-keeping, writing is now critical to communication and learning, as well as self-expression (Graham, 2006; Harris & Graham, 1992). Writing can foster a sense of heritage and purpose among members of communities or nations, help maintain connections with those important to us, enhance self-exploration, provide psychological and physiological benefits, act as a powerful tool in persuasion, and enhance the human experience in many additional ways (Graham, Harris, & Olinghouse, 2007; Swedlow, 1999).

Students who struggle significantly with writing, and adults who cannot or will not engage in writing, are at a terrible disadvantage in today's world. By the upper elementary grades, writing becomes a critical tool both for learning and for showing what you know. Writing is critical to gathering, refining, extending, preserving, and transmitting information and understandings; making ideas readily available for consideration, evaluation, and future discourse; fostering the examination of unexamined assumptions and creating cognitive disequilibrium that spurs learning; and promoting personal development (Graham, 2006; Prior, 2006). Students who do not write well cannot draw on its power to support and extend learning and development, and adults with inadequate writing skills can face significant barriers in further education and employment. Thus, challenges with writing can result in students failing to realize their educational, occupational, or personal potential.

Data from the National Assessment of Educational Progress, however, make it clear that many students in the United States find writing challenging; in both 1998 and 2002, the majority of fourth-, eighth- and twelfth-grade students demonstrated only partial mastery of the writing abilities needed at their grade level. Only 25% of students were classified as competent writers (Persky, Daane, & Jin, 2003). Five areas of competence have been identified as particularly difficult in learning to write: (a) generation of content, (b) creating an organizing structure for compositions, (c) formulation of goals and higher-level plans, (d) quickly and efficiently executing the mechanical aspects of writing, and (e) revising text and reformulating goals (Scardamalia & Bereiter, 1986). In addition, children in our schools frequently demonstrate a deteriorating attitude toward writing, even though most children begin school with a positive attitude toward composing.

Furthermore, almost one in every five first-year college students requires a remedial writing class, and more than one-half of new college students are unable to write a paper

relatively free of errors (Intersegmental Committee of the Academic Senates, 2002). National concern for the importance of teaching and learning writing led the College Board to establish the National Commission on Writing for America's Families, Schools, and Colleges in 2002. The Commission has since released several influential reports (found at www.writingcommission.org), including *The Neglected "R": The Need for a Writing Revolution* (2003). The need to keep writing at the center of school reform is emphasized in their more recent, fourth report to Congress, *Writing and School Reform* (2007).

In-depth understanding of the development of writing abilities, the difficulties that some students encounter and the reasons for these difficulties, and effective instructional practices are clearly essential. Research on writing, however, has occurred for only the past few decades, although the history of writing goes back over 500 years. Nystrand (2006) dated the launch of empirical research on writing as the early 1970s. Although some work had been done in the 1960s and earlier, these initial studies were isolated and of minimal impact. Virtually every author who writes about writing notes that it is a complex activity, and that learning to write is even more complex (cf. Sitko, 1998; McCutchen, 2006; Hayes, 2006). The writer must negotiate the rules and mechanics of writing, while maintaining a focus on factors such as organization, form and features, purposes and goals, audience perspectives and needs, and evaluation of communicative intent and efficacy (Bereiter & Scardamalia, 1987; Harris & Graham, 1992). Thus, it is no surprise that research on writing is also complex and challenging. Further, writing research has received far less funding and attention than research in reading or math, further challenging progress in this area.

In this chapter, our goal is to present critical information regarding metacognition and writing for both researchers and practitioners. First, we define metacognition and then present a broad overview of the progress we have made in writing research and understanding the role of metacognition in writing. How writing abilities, including metacognition necessary for writing, can be developed is then illustrated by reviewing the most extensively researched approach, strategies instruction.

Writing and Metacognition: Definition and Progress

Good writing does not reveal its making. The problems are solved, the scaffolding has been removed, and the discarded building materials have been hauled away.

D. M. Murray

An early model of the stages of writing was presented by Rohman in 1965, but this model posited three, linear stages in the writing process, and the notion that writing was a linear process was disputed in the 1970s (Zimmerman & Reiserberg, 1997). In 1971, Emig published a groundbreaking study of the composing processes of twelfth-graders. Her work and that of others began to establish the complex and reciprocal nature of the stages of writing. Empirical research in writing did not begin to have a major impact on education, however, until the 1980s. While other researchers also presented more complex models of writing, (e.g., Beaugrande, 1984), Hayes and Flower (1980) developed the model of writing that has been one of the two most cited and most influential models in the field of education. As Sitko (1998) noted, Hayes and Flower's model resembled the earlier work by Rohman, but advanced this early model by delineating the recursive nature of the writing process and by identifying sub-processes for each of the three major processes in writing: planning, translating, and reviewing. As this view of writing became more prominent, American educators increased their shift in attention from writing products to the writing process, and researchers and practitioners began developing instructional methods

for writing based on the process approach (for more detailed histories, see both Sitko, 1998 and Nystrand, 2006).

At the same time as changes were occurring in the way writing was conceptualized and researched, other important movements were happening. Both the new focus in writing research and a second movement, development of the construct of metacognition, were related to the cognitive revolution occurring in education and psychology. Metacognition became an important component of Hayes and Flower's (1980) model of writing, with metacognition playing a critical role in effective writing. The strongest impetus toward the exploration of this new construct came when Flavell (1971) noted that memory development appeared to include active, intelligent monitoring and knowledge of memory search and storage operations—what he termed “a kind of metamemory” (p. 277), and he enthusiastically encouraged researchers, “Lets all go out and study it!” Researchers in many areas, including writing, took up this call, and Flavell published a model of metacognition and cognitive monitoring based on research in 1979.

Space prevents a detailed history of our understanding of metacognition here (for more detail, see Hacker, 1998; McCormick, 2003; Pressley & Harris, 2006). As Hacker noted in 1998, the construct was “fuzzy” in many ways, and today researchers continue to refine it. In this chapter, we discuss the major components of metacognition generally accepted today. Metacognition has often been referred to as “cognition about cognition” (Brown, 1980). Two major, and commonly cited, elements of metacognition are: (a) knowledge about cognition and (b) deliberate, conscious regulation and control of cognitive activity (cf. Ertmer & Newby, 1996; Flavell, 1979; McCormick, 2003). It is, of course, more complicated than this.

Metacognition: Knowledge About Cognition

This element is typically extended to include both knowledge about cognition, or thinking, as well as awareness of one's own cognition. Awareness of one's own cognition is obviously necessary to both understand and control cognition. Researchers and theorists have identified three different kinds of metacognitive knowledge: declarative knowledge, procedural knowledge, and conditional knowledge (McCormick, 2003).

Declarative knowledge refers to one's knowledge about oneself as a learner, including knowledge about one's own abilities (strengths and weaknesses). Declarative knowledge also refers to knowledge regarding the task, including the knowledge, skills, and strategies needed for effectively completing the task under one or more conditions. In recent years, declarative knowledge has been expanded to include individuals' knowledge and understanding of their affective states, including self-efficacy and motivation, and how these characteristics affect task performance (cf. Hacker, 1998; Pressley & Harris, 2006; Zimmerman & Reiserberg, 1997). Thus, declarative knowledge encompasses knowledge about the self, the task, and strategies or procedures applicable to the task.

Procedural knowledge is the knowledge needed to carry out procedures, including strategies, in order to apply declarative knowledge and reach goals. This is knowledge about “how to do it.” Conditional knowledge refers to knowing when, where, and why to use declarative knowledge as well as particular procedures or strategies (procedural knowledge), and is critical to effective use of strategies. Effective performance among learners depends upon the application of declarative, procedural, and conditional knowledge. These three forms of knowledge vary among learners due to factors such as age, experience, interests, and so on. Without careful development of all three kinds of knowledge, success with writing is unlikely. Further, unless students experience initial success with writing, it is unlikely that they will expend the effort necessary to use more complicated

strategies and perform increasingly demanding writing tasks effectively (Pressley & Harris, 2006). Researchers and theorists continue to expand and refine the construct of metacognitive knowledge today.

Metacognition: Regulation and Control of Cognitive Activity

This aspect of metacognition has been addressed by numerous researchers. Paris and Lindauer (1982), applying this construct to both reading and writing, proposed that control of cognitive activity consists of planning, monitoring, and evaluating. Selection of strategies and allocation of resources occur during planning. Monitoring refers to analysis of the effectiveness of the strategies or plan being used. Evaluation refers to determining progress being made toward the goal, and can result in revisions or modifications to the initial plan (further planning), further monitoring, and further evaluation.

Researchers addressing the construct of metacognition, however, have not been the only group concerned with the active control of cognition or performance. While research in metacognition was progressing, so was research in two related areas: executive function and self-regulation (cf. McCormick, 2003; Harris, Graham, & Pressley, 1992). As McCormick noted, differing approaches to the concept of active control of cognition developed from different theoretical bases, assumptions, and methodologies. Over time, however, it is our observation that each of these approaches to understanding control of cognition have informed and impacted each other, making the distinctions among them less manifest. All of these constructs (metacognition, executive function, and self-regulation) have impacted both our knowledge of and instruction in writing, though we focus in this chapter on metacognition and self-regulation (cf. Ertmer & Newby, 1996; Hacker, 1998; Sitko, 1998; Zimmerman & Reisemberg, 1997). For further discussion of the distinctions between executive function and metacognition and their contributions to our understanding of learning and writing, see Meltzer (2007) and Graham et al. (2007).

Major Elements of Metacognition in Writing

Knowledge about cognition, the first major element of metacognition, includes declarative, procedural, and conditional knowledge. McCormick (2003) provided definitions of declarative metacognitive knowledge, procedural metacognitive knowledge, and conditional metacognitive knowledge in general terms. While we do not attempt to list exhaustively the specifics that each form of knowledge can take in relation to writing, we examine examples of each. In terms of writing, declarative knowledge can take many forms. First, there is the knowledge the writer has about themselves as a writer, including such things as what forms of writing have been engaged in successfully or unsuccessfully in the past, what components or elements of writing they are comfortable with and which they have not yet mastered (such as using dialogue in creative writing), and what environmental characteristics are preferable. In addition to knowledge about the self as a writer, there is knowledge regarding the writing task. This includes, but is not limited to, knowledge regarding mechanics, form, skills, and strategies (including strategies specific to a particular writing task, such as writing a persuasive essay, and general writing strategies such as an opening that catches the reader's interest, developing a voice, etc.) that are appropriate to the writing task at hand. Finally, declarative knowledge also includes the writer's knowledge about their own affect related to writing, including their self-efficacy for writing in general and the specific writing task (research indicates that for many struggling writers their self-efficacy for writing is not commensurate with their actual performance, with some students overestimating their abilities and others underestimating

themselves, cf. Harris et al., 1992), their motivation to write, and how these and other affective factors may influence their writing.

Procedural knowledge is needed in order to apply declarative knowledge and reach the writer's goals. This includes the general and genre-specific strategies the writer is knowledgeable of, as well as knowledge of how skills work and when they are needed. Procedural knowledge can also include knowledge about one's own optimal writing environment. Conditional knowledge allows the writer to determine when, where, and why to use their procedural and declarative knowledge. Conditional knowledge is involved as the writer evaluates the writing task and determines the skills and strategies needed, selects among alternative strategies, identifies the environmental conditions that can be addressed to make writing conducive (for example, one famous writer always rents a hotel room to begin a new novel), identifies when and why to engage in different components of the writing process, and so on.

The second major element of metacognition is the conscious regulation of writing activity. This is a complex element for skilled writers, as they seek to regulate such factors as selection and use of strategies and skills, the writing environment, managing cognitive load (including working and long-term memory, cf. Galbraith, Ford, Walker, & Ford, in press), their affective responses to writing, and attentional control (Graham et al., 2007; Hayes, 2006; Zimmerman & Reiserberg, 1997). Zimmerman and Reiserberg (1997) identified three classes of writer self-regulation, and listed 10 critical self-regulatory processes in writing (see Table 8.1). The three classes and 10 processes they identified are all important to consider in developing effective writing instruction, and thus have strongly influenced models of strategies instruction for writing.

Zimmerman and Reiserberg (1997), as noted earlier, also stressed the importance of self-efficacy, which is influenced by the success or failure of the self-regulatory strategies enacted, and which in turn influences motivation for writing and the further use of self-regulatory processes. Thus, writing interventions that consider the postulates of Zimmerman and Reiserberg will attend to the development of self-efficacy for writing, as

Table 8.1 Self-Regulation of Writing (Zimmerman and Reiserberg, 1997)

Classes and processes

- I. Environmental self-regulation. Processes: (1) environmental structuring (selecting, organizing, and creating effective writing settings) and (2) self-selected models that serve as sources of writing knowledge and skills. Many writers indicate the influence of models of writing at stages in their development. For example, they may seek to learn from an identified text or a particular writer's style.
 - II. Behavioral processes: (3) self-monitoring and self-recording one's own performance (i.e., charting the number of pages written, sections completed, or incremental goals met), (4) self-consequences for one's own performance (such as rewards for meeting goals or punishments for failing to meet goals), and (5) self-verbalizations used to enhance the process of writing (saying dialogue aloud as one composes or reading text aloud to assist in evaluation).
 - III. Personal and covert (unobservable) processes: (6) time planning and management (estimating and budgeting time needed for writing), (7) goal setting (specifying long- and short-term outcomes while writing, goals for the quality and characteristics of the written product, and so on), (8) setting self-evaluative standards for one's own writing (setting and adhering to specific standards of personal satisfaction regarding one's writing, such as setting criteria for judging an effective concluding paragraph), (9) use of cognitive strategies (rule governed methods for organizing, producing, and transforming written text, and (10) use of mental imagery to recall or create a vivid image of a setting, activity, or character to facilitate effective written description.
-

helping students develop self-efficacy for writing is critical in developing belief in themselves as writers and in the value of their writing.

Models of the Writing Process

Anyone who has composed a brilliant sentence and then forgotten the end of it before it could be committed to paper has experienced one of the problems that limited memory creates for writers.

J. R. Hayes

To further illustrate the role of metacognition in writing, we turn to important models of writing that guide much of research today. These models make the intricate and critical role of metacognition in writing evident, and help set the stage for our examination of metacognition and strategies instruction in writing. We briefly describe three models, returning first to the model developed by Hayes and Flower (1980, later revised by Hayes, 1996), and then to models developed by Bereiter and Scardamalia (1987), and Zimmerman and Reiserberg (1997). The major elements of metacognition in writing that we have identified, knowledge about writing and deliberate, conscious regulation and control of the writing process, are addressed in each model. Each model, however, contributes some unique perspectives.

Hayes and Flower; Hayes

Hayes and Flower's (1980) model of writing included three basic components: task environment, cognitive processes, and writer's long-term memory. Task environment included factors external to the writer, but that influenced the writing task, including attributes of the writing assignment and text produced so far. Cognitive processes referred to the mental operations employed during writing, and was broken into three important sub-processes: planning, translating, and reviewing. Further, the writer was seen as monitoring this complex process. In long-term memory, the writer held knowledge about the topic, the intended audience, the genre, and general plans or formulas for accomplishing varying writing tasks.

Hayes and Flower (1980) further divided the cognitive sub-process of planning into setting goals, generating ideas, and organizing ideas into a writing plan. Reviewing, similarly, was subdivided into evaluating and revising text. Importantly, Hayes and Flower postulated that the execution of these cognitive processes was under the writer's control, and that any process or sub-process could interrupt or incorporate any other sub-process, emphasizing the recursive, reciprocal nature of the writing process overall. Thus, in this model, as the writer develops further goals for their product while writing, planning will interrupt translation. As new ideas are generated while writing, the writing plan will require re-organization. Or, as McCormick (2003) noted, planning may be evoked during editing, or reviewing may serve to help organize. Many other interruptions or integrations are possible.

Hayes and Flower (1980) established several tenets that continue to impact research and instruction today. First, writing is a goal-directed activity (and thus, a problem-solving activity) that involves both major goals and subgoals. Second, effective writing requires a variety of mental operations necessary to reach the author's goals and subgoals. Third, the process of writing is not linear, but rather involves the complex interplay of interwoven or nested activities, and fourth, skilled writers must deal with a great many demands at once. These demands can include, but are not limited to, making plans, revising plans, drawing

knowledge and ideas from memory, developing concepts, imagining and responding to an uninformed or critical reader, otherwise considering reader needs, and managing the mechanics of writing (such as spelling, grammar, handwriting or keyboarding, and so on).

The model proposed by Hayes and Flower (1980) created a more uniform vocabulary for talking about and studying the writing process, and served as a catalyst for a great deal of research on the writing process and writing instruction. In 1996, Hayes made several modifications to the original model based on research and theory. First, as the vocabulary regarding writing had evolved, he changed the term translation to text generation, and the term reviewing to revising. In addition, in the revised model he postulated two major elements in the writing process, the task environment and the individual. He subdivided the task environment into the social environment and the physical environment. The social environment included the audience and collaborators, while the physical environment included both the text so far and the composing medium.

Hayes (1996) subdivided the second major element, the individual, into motivation/affect, cognitive processes, working memory, and long-term memory. Motivation/affect was subdivided into goals, predispositions, beliefs and attitudes, and cost/benefit estimates. Cognitive processes were subdivided into text interpretation, reflection, and text production. Working memory included phonological memory, visuospatial sketchpad, and semantic memory. Long-term memory included task schema, topic knowledge, audience knowledge, linguistic knowledge, and genre knowledge. While not all of the components of this revised model have been accepted into current vocabulary and understanding of the writing process, the model does reflect many advances in research and our increasingly complex understanding of the writing process. The original Hayes and Flower model (1980) and the revised model (Hayes, 1996) continue to have significant impact on writing research and instruction today.

Bereiter and Scardamalia

Bereiter and Scardamalia (1987) developed another model of writing that has had a major impact on research and instruction in writing today; their model and that of Hayes and Flower, and Hayes' revisions, are frequently cited as the two most influential models of writing in education and psychology (Graham, 2006). Bereiter and Scardamalia were interested in explaining critical differences between novice and expert writers, and in detailing the processes used by both. In their model, writing consisted of four main processes: (a) a mental representation of the task, (b) problem analysis and goal setting, (c) problem translation, and (d) resultant knowledge telling. Writing was described as a recursive problem-solving process, allowing expert writers to think and write more effectively on a topic. Further, two forms of strategies were used to improve writing: rhetorical and self-regulatory. Importantly, Bereiter and Scardamalia described novice writers as engaging in simple "knowledge telling," an understanding and term that is still used.

Knowledge telling among novice writers is seen as a greatly simplified version of the idea generation process in the Hayes and Flower model (1980; Graham, 2006). The knowledge telling model consists of three components. The first, mental representation of the assignment, involves understanding the assignment by defining the topic and function of the text to be produced. The second, long-term memory, includes two types of knowledge available to the writer, content knowledge (what the writer knows about the assigned topic) and discourse knowledge (linguistic knowledge and knowledge about the type of text to be produced). The third, the knowledge telling process, consists of operations writers use to produce text. The simple knowledge telling process used by novice writers basically involves retrieving content from long-term memory and writing it down.

In contrast, Bereiter and Scardamalia (1987) proposed a far different process used by more developed or expert writers, which they referred to as knowledge transforming. This complex process is important here, as it has often been the basis not only for research on writing, but has had a strong impact on the construction of instructional practices in writing that include a focus on developing metacognitive abilities, including strategies instruction. Thus, we detail it here. The knowledge transforming approach to writing involves planning text content in accordance with rhetorical (e.g., considerations of style, effect on reader), communicative (e.g., major points to be made, scope and breadth of content), and pragmatic (e.g., space, time for writing) constraints. Similar to the knowledge telling model, the starting point is to develop a mental representation of the assignment. Based on this understanding, the writer then engages in problem analysis and goal setting to determine what to say (content planning), as well as how to say it and who to say it to (rhetorical process planning). This is done by analyzing the task, setting content and rhetorical goals, and deciding on the necessary means to obtain these objectives. These two types of planning are carried out in their own space: content problem space and rhetorical problem space. Within these spaces, the writer retrieves and transforms knowledge about what they plan to say (content knowledge), as well as knowledge about their audience and how to say it (discourse knowledge).

These processes are guided by the goals and constraints established during problem analysis and goal setting. Planning in these two spaces operates in a close interaction through a problem translation component, which transfers goals and constraints from one space to the other. Thus, topic knowledge can be transformed by taking into account content goals as well as rhetorical and pragmatic constraints. Likewise, rhetorical and pragmatic dimensions can be transformed by content constraints. The resulting plans are then elaborated in writing through the knowledge telling process (described earlier). As the writer transcribes text using the knowledge telling process, the text is analyzed and the resulting information is fed back into the problem analysis and goal setting component, providing additional opportunities to engage in content and rhetorical planning based on the fit between intentions and the resulting written product. As can be seen here, the use of the knowledge telling process, and of the larger writing process, is quite different for novice versus expert writers.

Research by Bereiter and Scardamalia (1987), and many others, has resulted in important understandings of the differences between novice and expert writers (Graham, 2006; Sitko, 1998). Findings of this body of research indicate that immature and struggling writers use very simple processes in line with the simple version of the knowledge telling process proposed by Bereiter and Scardamalia. In the past three decades, a great deal of research has been done regarding the characteristics of writing among students with severe difficulties with writing, including students with learning disabilities (cf. MacArthur, Graham, & Fitzgerald, 2006; McCutchen, 2006). These students: typically produce writing that is less polished, expansive, coherent, and effective than that of their higher-achieving peers; lack critical knowledge of the writing process; have difficulty generating ideas and selecting topics; do little to no advance planning; engage in a simple form of knowledge telling; lack important strategies for planning, producing, organizing, and revising text; have difficulties with mechanics that interfere with the writing process; emphasize mechanics over content when making revisions; and frequently overestimate their writing abilities. Clearly, many of the difficulties encountered by novice and struggling writers involve metacognitive aspects of writing.

As we seek to apply these findings to instruction in writing and to move students toward more mature, expert-like writing processes, however, it is important to note that novice and expert writers represent two ends of a continuum, and that the development of writing

expertise does not appear to move from simple knowledge telling to knowledge transformation, but through several intermediate stages (Bereiter, Burtis, & Scardamalia, 1988). Further, while research currently supports the use of the simplified knowledge telling process by novice and struggling writers, there is as of yet insufficient evidence to indicate that the more complex knowledge transformation model is an accurate description of skilled writing (Graham, 2006).

Zimmerman and Reisemberg

A third model of writing is covered here, because it has important relevance to the topic of metacognition in writing. In 1997, Zimmerman and Reisemberg reviewed the models developed by Hayes and Flower and by Bereiter and Scardamalia. They argued that although these models included the task environment and self-regulatory strategies, they focused on the role of cognitive processes in students' writing *competence*, as opposed to writing *performance* and its self-regulated development. They further argued that "explanations focusing on writing performance and its self-regulated development need to include the role of social, motivational, and behavioral processes as well as cognitive ones" (p. 75). Working from social-cognitive theory and self-regulation theory (Zimmerman, 1989), they proposed a model of writing composed of three fundamental forms of self-regulation: environmental, behavioral, and covert or personal. The authors argued that these triadic forms of self-regulation interact reciprocally via a cyclic feedback loop that allows writers to self-monitor and self-react to feedback about the effectiveness of specific self-regulatory techniques or processes.

Zimmerman and Reisemberg (1997) argued that the self-regulation of writing involves a complex system of interdependent processes. Further, they placed particular emphasis on the construct of self-efficacy, as they proposed that these complex, interdependent processes are closely linked to an underlying sense of self-efficacy. A writer's sense of efficacy may be enhanced or diminished depending upon the perceived success of the self-regulatory strategies they put into play for controlling their actions, the writing environment, and their internal thoughts. Self-efficacy, in turn, influences intrinsic motivation for writing, the use of self-regulatory processes during writing, and eventual literary attainment.

It can be argued that by the time Zimmerman and Reisemberg (1997) published their model, many researchers and theorists from multiple theoretical perspectives, including cognitive psychology, had begun to strongly emphasize the importance of motivation, self-efficacy, and other affective constructs in writing, as can be seen in Hayes' (1996) revisions to the element of the individual in the earlier Hayes and Flower model. It can also be argued that Zimmerman and Reisemberg's model is more limited than others as it does not address the interaction between self-regulation and other processes involved in writing, such as working memory or text transcription skills. This model, however, provides an important contribution to descriptions of the composing process and our understanding of metacognition in writing for three reasons. First, Zimmerman and Reisemberg offer an explicit explanation of how writers exert deliberate control over the act of writing, via the 10 self-regulation processes in writing presented earlier. Even though writing is commonly viewed as a difficult and demanding task, requiring extensive self-regulation and attention control, the details of how writers use self-regulation strategies to manage the composing process received less detailed attention in previous models (Graham, 2006; Graham & Harris, 1997).

Second, Zimmerman and Reisemberg provided a compelling description of how writers' beliefs about competence influence and are influenced by their self-regulatory actions and

subsequent performance. This interplay was not explicitly examined in previous models. Finally, the model not only describes what writers do, but also addresses the processes of change as writers develop, by introducing mechanisms related to change, including thoughts, feelings, and actions, that writers use to attain differing literary goals. Zimmerman and Reisemberg did not limit writers' goals to enhancing the text that was produced, but also included goals for improving one's own performance as a writer.

Concluding Comments on Models of the Writing Process

Research in writing continues to expand and to impact teaching and learning. We close this section of this chapter by noting two ongoing issues in research relevant to metacognition in writing. The first is a definitional issue, and the second a theoretical one. Then, we briefly describe an important movement in the teaching of writing, as this impacts our discussion of the development of metacognition in writing through strategies instruction.

As noted earlier, deliberate, conscious regulation and control of cognitive activity is commonly seen as one of the major components of metacognition. Further, metacognition includes both strategies for addressing a task (such as genre-specific and general writing strategies) as well as self-regulation strategies (goal-setting, self-monitoring, self-instruction, self-reinforcement, etc.). The problem arises when the use of self-regulation strategies or writing strategies becomes automatized and therefore "non-conscious." Researchers and theorists today vigorously debate whether or not automatic (but potentially conscious if necessary) use of a writing or self-regulation strategy constitutes an occurrence of metacognition. In other words, can strategies operate at the realm of the unconscious (cf. Hacker, 1998; Harris, Alexander, & Graham, 2008; McCormick, 2003)? At issue is the distinction between typical or habitual cognitive behavior (i.e., skills) and more intentional or deliberate acts (i.e., use of strategies). We mention this issue here, because although it is still being debated, it does impact the definition of metacognition, our use of the term, and how we talk about instruction. For the purposes of this chapter, we will assume the deliberate, conscious use of a writing or self-regulation strategy when we discuss this aspect of metacognition.

Research on writing has been approached from many theoretical positions. So far in this chapter, we have referred to cognitive theory and social-cognitive theory as influential models for teaching writing and learning to write. Prior (2006), however, noted that sociocultural theories "represent the dominant paradigm for writing research today" (p. 54). Prior argued that Hayes and Flower's (1980) model has been critiqued as too narrow in its understanding and representation of context, and that current research from a sociocultural perspective attends far better to the social, historical, and political contexts of writing (see Nystrand, 2006, and Englert, Mariage, & Dunsmore, 2006 for further discussion). Further, sociocultural studies of writing have come from multiple disciplines, including anthropology, psychology, and education, and from multiple settings, such as the workplace, classrooms, homes, and communities.

Interestingly, as theories evolve, theorists respond to both critiques of their paradigm and to new ideas. Developments in sociocultural theory have occurred in the past few decades alongside developments in cognitive, information processing, social-cognitive theories, and others. Thus, over time it can become more difficult to ascertain critical differences between some theories, especially at an applied level. In this chapter, we do not assume any single theoretical perspective. Rather, as Harris and Graham, as well as others, have argued (cf. Harris, 1982; Harris & Graham, 1992), we take the perspective that multiple theories can inform powerful interventions needed in our schools today. This position might be referred to as theoretical pragmatism. The challenges faced by students

with special needs, and indeed by all of us today, are complex. When we treat competing viewpoints with thoughtfulness and respect, a powerful repertoire for teaching and learning can be developed. This does not negate the important contributions made by competing theories, and the importance of this competition in advancing our thinking and research (Harris, Graham, & Mason, 2003; Harris & Graham, 1992; Harris & Alexander, 1998).

The Process Approach to Writing Instruction

As research on, and understanding of, the writing process developed since the 1960s, the predominant approach to writing instruction in this country switched from a product approach (mechanics and grammar tended to be emphasized over content and process) to a process approach. In the process approach, teachers create an environment where students have time not only to write, but to think and reflect upon what they are writing about. Instruction takes place in a supportive environment where students are encouraged to choose their own topics, help each other, and take risks. Students write for real purposes and for real audiences, rather than just for their teachers, and are given opportunities for extended writing.

An important goal in the process approach is for students to see writing as a process, albeit one that is difficult and frustrating at times, yet also a challenging and enjoyable vehicle for learning and self-expression (cf. Atwell, 1987; Graves, 1985). While implementation of the process approach to writing has taken many forms and in some cases created considerable controversy (Harris & Graham, 1994), it has been an important step forward in responding to the calls to improve writing instruction in this country. Further, the emphasis on process rather than product alone clearly addressed the development of metacognition in writing.

Strategies Instruction in Writing

Writing is easy. All you have to do is sit down at a typewriter and open a vein.

R. Smith

We focus on research on strategies instruction in regards to metacognition and writing instruction in this chapter for two reasons. First, strategies instruction interventions have clearly addressed multiple metacognitive components of writing. Second, strategies instruction has received a great deal of research attention and has had the strongest impact on writing performance among school-age students of any intervention researched. Drawing on three sources that help synthesize writing research to date, including a meta-analysis of experimental and quasi-experimental writing intervention studies (Graham & Perin, 2007a, 2007b), a meta-analysis of single-subject design studies (Rogers & Graham, 2007), and a meta-synthesis of qualitative studies examining writing instructional practices among effective teachers and schools (Graham & Perin, 2007c), Graham and Harris (in press) listed 12 evidence-based recommendations for improving writing among students in grades 4 to 12.

These evidence-based recommendations clearly illustrate the importance of metacognition in writing. Four of these recommendations were supported by all three sources, in this order (based on criteria regarding the strength of the evidence and impact on writing): (a) teach strategies for planning, revising, and editing writing, (b) set clear and specific goals for what writers are to accomplish in their writing product, (c) help writers learn to write more sophisticated sentences, and (d) engage students in prewriting activities that

help them gather and organize ideas for their compositions. Strategies instruction explicitly incorporates three of these four; writing more sophisticated sentences has not yet been a major element in strategies instruction, though it clearly should be in future research.

The next eight recommendations, ordered by the same criteria, are: (e) engage students in the process writing approach, (f) teach students strategies and procedures for summarizing reading material, as this improves their ability to concisely and accurately present this information in writing, (g) incorporate instructional arrangements that allow students to work together to plan, draft, revise, and edit their compositions, (h) make it possible for students to use word processing as a tool for writing, (i) involve students in writing activities designed to sharpen their inquiry skills, (j) provide good models for each type of writing that is the focus of instruction, (k) have students monitor their writing performance or behavior, and (l) provide ample time for writing (Graham and Harris, in press). Strategies instruction research has addressed many, but not all of these (see Wong, Harris, Graham, & Butler, 2003, for more details). Some researchers have begun development of strategies instruction that addresses both reading and writing (e.g., developing strategies for reading and writing expository text in tandem) and others have begun investigations of effective peer support for writing (Pressley & Harris, 2006). Further, strategies instruction typically includes monitoring of performance, good models of writing, and providing time for writing. Further research, however, is clearly needed in all of these areas, and in incorporating inquiry with strategies instruction.

Approaches to Strategies Instruction in Writing

Harris, Graham and their colleagues, and Englert and her colleagues, have developed strategy instruction models for composition. Harris and Graham's approach has come to be called Self-Regulated Strategy Development (SRSD); the first study using this model was published in 1985. Englert and her colleagues refer to their approach as the Cognitive Strategy Instruction for Writing (CSIW) Program and published two influential studies involving CSIW. Results of these studies indicated that both students with and without learning disabilities (LD) improved their knowledge of the writing process and their writing abilities. In the Englert et al. (1991) study, students with LD performed similarly to their peers without disabilities on all five post-test writing variables. In the 1992 study, Englert, Raphael, and Anderson found that the quality of students' metacognitive knowledge was positively related to measures of performance in both reading and writing.

Since 1985, more than 40 studies using the SRSD model of instruction have been reported in the area of writing, involving students from the elementary grades through high school (Graham & Harris, 2003, in press). In the meta-analyses of true- and quasi-experimental design studies, SRSD has had the strongest impact of any strategies instruction approach in writing. Thus, in the space we have here, we turn to a discussion of this model.

SRSD and Children's Writing

Inspiration is wonderful when it happens, but the writer must develop an approach for the rest of the time . . . the wait is simply too long.

L. Bernstein

In this section, we begin with a summary of the research findings regarding SRSD. Once we have established the research base for SRSD and its use in schools, we then provide a

description of the development of SRSD and the major components of SRSD instruction in writing. Finally, we note further research needs in this area.

Effectiveness of SRSD

In literature searches conducted to identify effective writing practices (Graham & Perin, 2007a, 2007b, 2007c; Rogers & Graham, 2007), 15 true- and quasi-experimental studies and 27 single-subject design investigations examining the effectiveness of SRSD writing interventions were identified. The overall impact of SRSD in the 27 single-subject design studies was reported in Rogers and Graham (2007) and is summarized here. Although several meta-analyses have included a majority of the 15 true- and quasi-experimental design investigations (Graham, 2006; Graham & Harris, 2003; Graham & Perin, 2007a), there is no current review that includes all of them. We rectify that situation here.

True- and Quasi-Experimental Design SRSD Writing Studies

We start first with the true- and quasi-experimental design studies. Eight of these 15 investigations were conducted by Harris and Graham, or in conjunction with one or both of them (De La Paz & Graham, 1997, 2002; Graham, Harris, & Mason, 2005; Harris, Graham, & Mason, 2006; Harris, Graham, & Adkins, 2004; MacArthur, Schwartz, & Graham, 1991; Sawyer, Graham, & Harris, 1992; Tracy, Reid, & Graham, 2007), whereas the other seven studies were not (Anderson, 1997; Curry, 1997; De La Paz, 2005; Garcia-Sanchez & Fidalgo-Redondo, 2006; Glaser & Brunstein, 2007; Torrance, Fidalgo, & Garcia, 2007; Wong, Hoskyn, Jai, Ellis, & Watson, 2007). Most of these studies involved the teaching of strategies for planning and drafting compositions in specific genres (e.g., narrative, persuasive, and expository), but some focused on a range of writing processes (e.g., Torrance et al., 2007) or just revising/editing (MacArthur et al., 1991). Students in these studies were in grades 2 to 8. In nine of the studies, students were poor writers (including children with learning disabilities), whereas students representing the full range of writing abilities in a typical class were included in seven studies (one study included both groups; Anderson, 1997). In all 15 studies, the impact of SRSD on writing quality was assessed. We focus our meta-analysis in this paper on this variable.

Effect sizes were calculated as Cohen's *d* or the standardized mean difference for all of the true- and quasi-experimental studies. An effect size for writing quality was computed for all 15 studies for the writing probe administered immediately following treatment (post-test). In calculating a mean effect size for all 15 studies, each effect size was weighted according to sample size. This assumes that studies with more participants provide better estimates of population parameter than studies with fewer participants. Although it is best to interpret the magnitude of an effect size in relation to the distribution of other mean effect sizes in the same general area, a widely used rule of thumb is that an effect size of 0.20 is small, 0.50 is medium, and 0.80 is large (Lipsey & Wilson, 2001).

The average weighted effect size for writing quality at post-test across the 15 SRSD studies was 1.20 (effect sizes ranged from 0.39 to 3.32). The confidence interval surrounding this average weighted effect size was 1.05 to 1.34. Thus, SRSD had a strong impact on improving the writing quality of the students participating in these studies. Importantly, there was no statistical difference between the average weighted effect size for studies with just poor writers and those with the full range of writing ability. Nor was there a statistically significant difference between studies involving primary grade, intermediate grade, and middle-school students.

In 9 of the 15 studies, a maintenance writing probe was administered anywhere from

1 week to 2 years following instruction (four studies were with poor writers, four were with writers representing the full range of classroom abilities, and one was with both; students were in grades 2 to 7). In one instance, a maintenance probe was administered 12 weeks following treatment (reported in Torrance et al., 2007; effect size = 2.75) and another maintenance probe was given approximately 2 years later (Fidalgo, Torrance, & Garcia, 2008; effect size = 0.78). In the case of this investigation, we obtained an effect size by averaging Cohen's d for the probes at these two time points. The average weighted effect size at maintenance for writing quality was large (1.23), and all maintenance effect sizes were positive (ranging from 0.65 to 1.77). The confidence interval surrounding this average weighted effect size was 1.04 to 1.42.

Finally, five of the true- and quasi-experimental studies examined if the effects of SRSD generalized to an untaught genre (three studies involved poor writers, one study involved students representing the typical range of classroom writing ability, and one study involved both). These five studies only involved students in grades 2, 3, and 5. The average weighted effect size for writing quality on the generalization probes was large (0.80), and all generalization effect sizes were positive (ranging from 0.57 to 1.58). The confidence interval surrounding this average weighted effect size was 0.57 to 1.03.

Single-Subject Design SRSD Writing Studies

The findings for the 27 single-subject SRSD writing studies reported by Rogers and Graham (2007) mirror those presented above. In this meta-analysis, the most commonly used single-subject effect size metric, percent of non-overlapping data (PND), was used (Scruggs & Mastropieri, 2001). PND is the percentage of non-overlapping data for a given treatment condition that exceeds the most positive value obtained during baseline (see Horner et al., 2005, for a description of single-subject designs and their logic). With single-subject design studies, PND can only be calculated for measures where the score for each data point is provided during baseline and treatment (usually in the form of a graph). The measure that is most likely to be graphed is the one that is most directly associated with the intervention (this is not typically a measure of writing quality). The measure that was most commonly graphed in the SRSD single-subject design studies was number of schematic elements (i.e., basic parts of a composition; such as characters, location, actions, ending, and so forth for stories), but length and quality were graphed in enough studies that Rogers and Graham computed a summary PND for these as well (no correction for number of participants in a study were made when computing the summary PND, as such statistical procedures are not currently available for single-subject design research).

Average PND of 90% or greater is considered strong, 70–90% moderate, and 50–70% small (Scruggs & Mastropieri, 2001). It must be noted that PND does not indicate the magnitude of an effect, as it reflects the percentage of treatment data points that represent an improvement over the strongest baseline score. Rather, PND provides an indication of how confident one can be that the treatment does have an effect. In addition, PND for a single study is typically based on a relatively small number of participants (e.g., 3 or 4 students).

Rogers and Graham (2007) located 25 single-subject design studies where students in grades 2 through 10 were taught strategies for planning and drafting either narrative, expository, or persuasive text (two additional studies taught revising/editing strategies). Most of these studies involved poor writers. In 21 of these 25 studies, the number of schematic elements was graphed. The mean PND for schematic structure at post-test for these studies was 97% (strong PND). In all but 3 of these 21 studies, schematic structure maintenance effects were also graphed. Again, the mean PND was strong (90%). A

small set of studies (four) assessed if SRSD effects generalized to an uninstructed genre. The mean PND was moderate (85%).

Ten of the 25 planning and drafting SRSD studies graphed writing output (number of words written) at post-test. Similar to schematic structure, a strong mean PND was obtained (91%). Six of these 10 studies graphed maintenance data for writing output. Mean PND was moderate (84%). Only 5 of the 25 planning and drafting SRSD investigations graphed writing quality, but mean PND for this small set of studies was strong (99%).

Summary

Data from a relatively large number of large-group and single-subject design studies provide convincing evidence SRSD is an effective method for teaching writing strategies to youngsters (both poor writers and students who represent the full range of writing ability in a typical class). SRSD research has resulted in the development of writing strategies (typically with the assistance of teachers and their students) for a variety of genres; these include personal narratives, story writing, persuasive essays, report writing, expository essays, and state writing tests. SRSD has resulted in significant and meaningful improvements in children's development of planning and revising strategies, including brainstorming, self-monitoring, reading for information and semantic webbing, generating and organizing writing content, advanced planning and dictation, revising with peers, and revising for both substance and mechanics (cf. Harris, Graham, & Mason, 2003).

SRSD has resulted in improvements in five main aspects of students' performance: genre elements included in writing, quality of writing, knowledge of writing, approach to writing, and self-efficacy. Across a variety of strategies and genres, the quality, length, and structure of students' compositions have improved. Depending on the strategy taught, improvements have been documented in planning, revising, content, and mechanics (Wong et al., 2003). These improvements have been consistently maintained for the majority of students over time, with some students needing booster sessions for long-term maintenance, and students have shown generalization across settings, persons, and writing media.

Development of SRSD

Harris and Graham began development of the SRSD approach to writing instruction with the underlying premise that students who face significant and often debilitating difficulties would benefit from an approach to instruction that deliberately and directly addressed their affective, behavioral, and cognitive characteristics, strengths, and needs (Harris, 1982). Further, these students often require more extensive, structured, and explicit instruction to develop skills, strategies (including academic, social, and self-regulation strategies), and understandings that their peers form more easily. The level of explicitness of instruction, however, should be adjusted to meet student needs (Harris & Graham, 1992). This perspective requires that the same academic and self-regulation strategies are not necessarily targeted for all students, and that instructional components and processes need to be individualized. As students' learning and behavioral challenges become more significant, strategy and self-regulation development becomes more complex and explicit, involving multiple learning tasks, components, and stages.

The SRSD approach to strategies instruction views learning as a complex process that relies on changes that occur in learners' skills and abilities, including many aspects of metacognition: self-regulation, strategic knowledge, domain-specific knowledge and abilities, self-efficacy, and motivation. There exists to date no single theory of teaching or learning that addresses all of the challenges faced by struggling learners; no current single

theories, in fact, fully capture complex phenomena such as learning (Harris, 1982). Thus, a further premise evident from the beginning of Harris and Graham's work on SRSD was the need to integrate multiple lines of research from multiple theoretical perspectives in order to develop powerful interventions for students who face significant academic challenges.

Thus, the development of SRSD has been, and continues to be, informed by research based on multiple theoretical perspectives, including behavioral, information processing, cognitive, social cognitive, constructivist, sociocultural, and other theories. Multiple bodies of research resulting from these perspectives have impacted and continue to impact development of SRSD, including research on motivation, written language, writing instruction, self-regulation, strategies instruction, expertise, learning characteristics of students with significant learning problems, and effective teaching and learning (see Harris & Graham, 1992, and Harris, Graham, & Mason, 2006, for further discussion).

The SRSD Approach to Strategies Instruction in Writing

Here, we present a brief overview of SRSD instruction. Detailed descriptions of SRSD instruction are available, however (see Harris, Graham, Mason, & Friedlander, 2008; Harris, Graham, & Mason, 2003). Detailed lesson plans and support materials for instruction are provided in Harris et al., 2008. All of the stages of instruction can be seen in both elementary and middle school classrooms in the video, "Teaching students with learning disabilities: Using learning strategies" (Association for Supervision and Curriculum Development, 2002). Finally, online interactive tutorials on SRSD are available at: <http://iris.peabody.vanderbilt.edu/index.html>. The tutorials include all stages of instruction and video from the ASCD video. From the IRIS homepage, select Resources, and then select Star Legacy Modules. For the first tutorial, click on "Using Learning Strategies: Instruction to Enhance Learning." For a tutorial on a persuasive writing strategy for elementary students, locate the header "Differentiated Instruction," then click on the module titled "Improving Writing Performance: A Strategy for Writing Expository Essays." Finally, a website devoted to strategies instruction can be found at <http://www.unl.edu/csi/>.

Critical Characteristics

There are five critical characteristics of SRSD instruction. One, writing (genre-specific and general) strategies and self-regulation strategies, as well as declarative, procedural, and conditional knowledge are explicitly taught and supported in development. Two, children are viewed as active collaborators who work with the teacher and each other during instruction. Three, instruction is individualized so that the processes, skills, and knowledge targeted for instruction are tailored to children's needs and capabilities. Goals are adjusted to current performance for each student, with more capable writers addressing more advanced goals. Instruction is further individualized through the use of individually tailored feedback and support. Four, instruction is criterion-based rather than time-based; students move through the instructional process at their own pace and do not proceed to later stages of instruction until they have met criteria for doing so. Importantly, instruction does not end until the student can use the strategy and self-regulation procedures efficiently and effectively. Five, SRSD is an ongoing process in which new strategies are introduced and previously taught strategies are upgraded over time.

SRSD has been used successfully with entire classes, small groups, and in tutoring settings (Graham & Harris, 2003). Classroom teachers are as successful, or more successful, than research assistants in implementing SRSD (Graham & Harris, 2003; Harris, Graham,

& Mason, 2006). Finally, lessons have typically run anywhere from 20 to 40 minutes (depending on grade level and class schedules) three to five days a week. In most of our work with teachers and students, instruction takes less time than teachers anticipated. In the elementary grades, eight to twelve 30–40 minute lessons conducted over 3–5 weeks have typically been what students need to reach independent use of the writing and self-regulation strategies (further details by grade and genre can be found in Graham & Harris, 2003).

The Process of Instruction

Six basic stages of instruction are used to introduce and develop genre-specific and general writing and self-regulation strategies in the SRSD approach. These stages are outlined in greater detail in Table 8.2. Throughout the stages, teachers and students collaborate on the acquisition, implementation, evaluation, and modification of these strategies. The stages are not meant to be followed in a “cookbook” fashion. Rather, they provide a general format and guidelines. The stages can be reordered, combined (in fact, most lessons include at least two stages, as can be seen in the lesson plans available noted earlier), revisited, modified, or deleted to meet student and teacher needs. Further, the stages are meant to be recursive—if a concept or component is not mastered at a certain stage, students and teachers can revisit or continue aspects of that stage as they move on to others. Some stages may not be needed by all students. For example, some students may already have the knowledge needed to use the writing and self-regulation strategies, and may skip this stage or act as a resource for other students who need this stage.

Metacognition and Strategies Instruction

The aspects of metacognition in writing we have discussed can clearly be seen in Table 8.2. Application of what we have learned from both models of writing and research on novice, struggling, and better writers is also evident in the SRSD approach. In SRSD students read model pieces and discuss these pieces, including the author’s goals when writing, the writing and self-regulation strategies the author may have used, and identification of genre elements and characteristics of good writing used by the author. The strategies instruction they are receiving is also explained and discussed to help develop declarative, procedural, and conditional knowledge. Writing is treated as a problem-solving process, and students set and monitor goals for their compositions. Self-regulation strategies are used to implement and coordinate writing strategies and to evaluate performance. In order to support both working and long-term memory, students are at first provided with a variety of prompts, including charts for the writing strategies as well as graphic organizers for planning. These prompts are slowly withdrawn until the student has memorized the mnemonics involved in their use and what each step of the mnemonic means—and students are given time and practice to memorize these mnemonics and their meanings. Students chart their progress as they learn the strategies, and attributions to effort and strategies use are emphasized and modeled. Beliefs about writing and the self as a writer that can help or hinder us as we write are discussed. Increasing self-efficacy in writing is a goal, and is supported by many elements of strategies instruction as seen in Table 8.2. Improving self-efficacy is addressed through a focus on attributions to effort and strategy use; student commitment to learning and using strategies; use of goal-setting, self-monitoring, and self-reinforcement; discussion of beliefs about writing; and scaffolding of initial writing performance. Finally, students are given the time they need to learn and develop initial competence in the use of the writing and self-regulation strategies. That this combination

Table 8.2 SRSD Stages of Instruction

Develop and activate knowledge needed for writing and self-regulation

- read works in the genre being addressed (stories, persuasive essays, etc.), to develop declarative, procedural, and conditional knowledge (e.g., what is an *opinion*?, what are the parts of a persuasive essay, are they all here?, how do you think the author came up with this idea?, what would you do?, what might the author have done to help herself come up with all of these ideas?, what might the author have done to organize the ideas?, What might the author do when he gets frustrated?, and so on), appreciation of characteristics of effective writing (how did the writer grab your interest?), and other knowledge and understandings targeted for instruction. Continue development through the next two stages as needed until all key knowledge and understandings are clear
- discuss and explore both writing and self-regulation strategies to be learned; may begin development of self-regulation, introducing goal setting and self-monitoring

Discuss it

- explore students' current writing and self-regulation abilities, their attitudes and beliefs about writing, what they are saying to themselves as they write, and how these might help or hinder them as writers
- graphing (self-monitoring) may be introduced, using prior compositions; this may assist with goal setting; graphing prior writing can be skipped if the student is likely to react negatively and only performance during instruction graphed
- further discuss strategies to be learned: purpose, benefits, how and when they can be used or might be inappropriate (begin generalization support)
- establish student commitment to learn strategy and act as collaborative partner; establish role of student effort and strategy use

Model it

- teacher modeling and/or collaborative modeling of writing and self-regulation strategies, resulting in appropriate model compositions
- analyze and discuss strategies and model's performance; make changes as needed
- can model self-assessment and self-recording through graphing of model compositions
- continue student development of self-regulation strategies across composition and other tasks and situations; discuss use here and in other settings (continue generalization support)

Memorize it

- though typically begun in earlier stages, require and confirm memorization of strategies, mnemonic(s), and self-instructions as appropriate
- continue to confirm and support memorization in following stages, make sure students have memorized the mnemonics and what they mean before independent performance

Support it

- teachers and students use writing and self-regulation strategies collaboratively to achieve success in composing, using prompts such as strategy charts, self-instruction sheets, and graphic organizers (can initially use pictures, then fade pictures in graphic organizers)
- challenging initial goals for genre elements and characteristics of writing established collaboratively with individual students; criterion levels increased gradually until final goals met
- prompts, guidance, and collaboration faded individually (graphic organizer replaced with student creating mnemonic on scratch paper) until the student can compose successfully alone
- self-regulation components not yet introduced may begin (typically, goal setting, self-instructions, self-monitoring and self-reinforcement are all being used by this stage; additional forms of self-regulation, such as environmental control, use of imagery, and so on may be used as desirable)
- discuss plans for maintenance, continue support of generalization

Independent performance

- students able to use writing and self-regulation strategies independently; teachers monitor and support as necessary
 - fading of overt self-regulation may begin (graphing may be discontinued)
 - plans for maintenance and generalization continue to be discussed and implemented
-

of components is effective can be seen in the data we have shared, and in the comment of one third-grader, who said to the first author, "Well of course now I can write, somebody taught me how!"

Generalization and Maintenance

Procedures for promoting maintenance and generalization are integrated throughout the stages of instruction in the SRSD model. These include: identifying opportunities to use the writing and/or self-regulation strategies in other classes or settings, discussing attempts to use the strategies at other times, reminding students to use the strategies at appropriate times, analyzing how these processes might need to be modified with other tasks and in new settings, and evaluating the success of these processes during and after instruction. It is helpful to involve others, including other teachers, parents, and other professionals (some teachers have involved speech language professionals, for example) as they can prompt, support, and evaluate the use of the strategies at appropriate times in other settings. Booster sessions, where the strategies are reviewed and discussed and supported again if necessary, are very important for some of the students we have worked with in terms of maintaining the strategies.

Research Needs

Despite the relatively large body of research on SRSD, many issues remain to be addressed. One of the most intriguing questions is the long-term results of strategies instruction and development of self-regulation across the grades in writing, and in other domains. No such research has been conducted; the longest studies have involved teaching two genre-specific writing strategies (such as story writing and persuasive writing) as well as appropriate general writing strategies within a single school year (Graham & Harris, 2003). At this time, we do not know how many genre specific strategies can most effectively be taught in a school year across different ages, or how these strategies can be enhanced over the years (cf. Harris & Graham, 1996). Parents and others could be partners in such long-term intervention, and research is needed here. Long-term research such as this is expensive, and funding is needed to support such efforts. Researchers have also argued that a focus on how teachers become adept at, committed to, and supported in strategy instruction is needed, as is more work aimed at filtering this approach into the schools. First, teachers must become knowledgeable of research supported practices; then they must decide if their classroom and students are an appropriate match to the treatment and validating data; and finally, they must implement and evaluate the effects of the treatment with their own students (Graham, Harris, & Zito, 2005). Further, adopting research-validated interventions can be very challenging to teachers because it often means teaching in new ways, with teachers typically not adequately supported in terms of professional development.

Harris and Graham have emphasized that SRSD should not be thought of as a panacea; promoting students' academic competence and literacy requires a complex integration of skills, strategies, processes, and attributes (Harris, 1982; Harris & Graham, 1996; Harris, Graham, & Mason, 2003). While SRSD represents an important contribution to teachers' instructional repertoires, it does not represent a complete writing curriculum. No research to date has been conducted examining the contribution of SRSD to the larger writing curriculum; such research is clearly needed. While SRSD helps get children "on the playing field" for writing, research has not gone on to address how to take this beginning and continue children's development of expertise in writing. Among skilled writers, writing is a flexible, goal-directed activity that is scaffolded by a rich knowledge of cognitive processes

and strategies for planning, text production, and revision. Skilled writers engage in purposeful and active self-direction of these processes and strategies (Harris, Schmidt, & Graham, 1998). We have a long road to go in determining how to bring children who struggle with writing to this level.

References

- Anderson, A. A. (1997). *The effects of sociocognitive writing strategy instruction on the writing achievement and writing self-efficacy of students with disabilities and typical achievement in an urban elementary school*. Unpublished doctoral dissertation, University of Houston, Houston, TX.
- Association for Supervision and Curriculum Development (2002). *Teaching students with learning disabilities in the regular classroom: Using learning strategies* [videotape 2]. Retrieved March 1, 2009 from <http://shop.ascd.org/productdisplay.cfm?productid=602084>
- Atwell, N. (1987). *In the middle: Reading, writing, and learning from adolescents*. Portsmouth, NH: Heinemann.
- Beaugrande, R. de (1984). Text production: Toward a science of composition. Norwood, NJ: Ablex.
- Bereiter, C., Burtis, P., Scardamalia, M. (1988). Cognitive operations in constructing main point in written composition. *Journal of Memory and Language*, 27, 261–278.
- Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.
- Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453–481). Hillsdale, NJ: Erlbaum.
- Curry, K. A. (1997). *A comparison of the writing products of students with learning disabilities in inclusive and resource room settings using different writing instruction approaches*. Unpublished doctoral dissertation, Florida Atlantic University, Boca Raton, FL.
- De La Paz, S. (2005). Teaching historical reasoning and argumentative writing in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology*, 97, 139–158.
- De La Paz, S. & Graham, S. (1997). Effects of dictation and advanced planning instruction on the composing of students with writing and learning problems. *Journal of Educational Psychology*, 89, 203–222.
- De La Paz, S. & Graham, S. (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology*, 94, 291–304.
- Emig, J. (1971). *The composing processes of twelfth graders*. Urbana, IL: National Council of Teachers of English.
- Englert, C. S., Mariage, T., & Dunsmore, K. (2006). Tenets of sociocultural theory in writing instruction research. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 208–221). New York: Guilford.
- Englert, C. S., Raphael, T. E., & Anderson, L. M. (1992). Socially mediated instruction: Improving students' knowledge and talk about writing. *Elementary School Journal*, 92, 411–449.
- Englert, C. S., Raphael, T. E., Anderson, L., Anthony, H., Stevens, D., & Fear, K. (1991). Making writing strategies and self-talk visible: Cognitive strategy instruction in writing in regular and special education classrooms. *American Educational Research Journal*, 28, 337–373.
- Ertmer, P. A. & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, 24, 1–24.
- Fidalgo, R., Torrance, M., & Garcia, J. (2008). The long term effects of strategy-focused writing instruction for grade six students. *Contemporary Educational Psychology*, 33, 672–693.
- Flavell, J. H. (1971). What is memory development the development of? *Human Development*, 14, 272–278.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906–911.
- Galbraith, D., Ford, S., Walker, G., & Ford, J. (in press). The contribution of different components

- of working memory to knowledge transformation during writing. *L1-Educational Studies in Language and Literature*.
- Garcia-Sanchez, J. & Fidalgo-Redondo, R. (2006). Effects of two types of self-regulatory instruction programs on students with learning disabilities in writing products, processes, and self-efficacy. *Learning Disability Quarterly*, 29, 181–211.
- Glaser, C. & Brunstein, J. (2007). Improving fourth-grade students' composition skills: Effects of strategy instruction and self-regulatory procedures. *Journal of Educational Psychology*, 99, 297–310.
- Graham, S. (2006). Writing. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (pp. 457–478). Mahwah, NJ: Erlbaum.
- Graham, S. & Harris, K. R. (1997). Self-regulation and writing: Where do we go from here? *Contemporary Educational Psychology*, 22, 102–114.
- Graham, S. & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 323–344). New York: Guilford.
- Graham, S. & Harris, K. R. (in press). Evidence-based writing practices: Drawing recommendations from multiple sources. *British Journal of Educational Psychology*.
- Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge, and motivation of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology*, 30, 207–241.
- Graham, S., Harris, K. R., & Olinghouse, N. (2007). Addressing executive function difficulties in writing: An example from the Self-Regulated Strategy Development model. In L. Meltzer (Ed.), *Executive functioning in education: From theory to practice* (pp. 216–236). New York: Guilford.
- Graham, S., Harris, K. R., & Zito, J. (2005). Promoting internal and external validity: A synergism of laboratory experiments and classroom based research. In G. Phye, D. H. Robinson, & J. Levin (Eds.), *Experimental methods for educational intervention* (pp. 235–265). San Diego, CA: Elsevier.
- Graham, S. & Perin, D. (2007a). *Writing next: Effective strategies to improve writing of adolescent middle and high school*. Alliance for Excellence in Education. Washington, DC.
- Graham, S. & Perin, D. (2007b). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, 445–476.
- Graham, S. & Perin, D. (2007c). What we know, what we still need to know: Teaching adolescents to write. *Scientific Studies in Reading*, 11, 313–336.
- Graves, D. (1985). All children can write. *Learning Disability Focus*, 1, 36–43.
- Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1–23). Mahwah, NJ: Erlbaum.
- Harris, K. R. (1982). Cognitive-behavior modification: Application with exceptional students. *Focus on Exceptional Children*, 15 (2), 1–16.
- Harris, K. R. & Alexander, P. A. (1998). Integrated, constructivist education: Challenge and reality. *Educational Psychology Review*, 10(2), 115–127.
- Harris, K. R., Alexander, P., & Graham, S. (2008). Michael Pressley's contributions to the history and future of strategy research. *Educational Psychology*, 43, 86–96.
- Harris, K. R. & Graham, S. (1992). Self-regulated strategy development: A part of the writing process. In M. Pressley, K. R. Harris, & J. Guthrie (Eds.), *Promoting academic competence and literacy in school* (pp. 277–309). New York: Academic Press.
- Harris, K. R. & Graham, S. (1994). Constructivism: Principles, paradigms, and integration. *The Journal of Special Education*, 28, 233–247.
- Harris, K. R., & Graham, S. (2007). "Marconi invented the television so people who couldn't afford radio could hear the news": The research on powerful composition strategies we have, and the research we need. In M. Pressley (Ed.), *Literacy achievement research we have, literacy research we need*. New York: Guilford.
- Harris, K. R., Graham, S., & Adkins, M. (2004, February). The effects of teacher-led SRSD

- instruction on the writing performance of struggling writers. Presentation at Pacific Coast Research Conference, CA.
- Harris, K. R., Graham, S., & Mason, L. (2003). Self-regulated strategy development in the classroom: Part of a balanced approach to writing instruction for students with disabilities. *Focus on Exceptional Children*, 35, 1–16.
- Harris, K. R., Graham, S., & Mason, L. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal*, 43, 295–340.
- Harris, K. R., Graham, S., Mason, L., & Friedlander, B. (2008). *Powerful writing strategies for all students*. Baltimore: Brookes.
- Harris, K. R., Graham, S., & Pressley, M. (1992). Cognitive-behavioral approaches in reading and written language: Developing self-regulated learners. In N. N. Singh & I. L. Beale (Eds.), *Learning disabilities: Nature, theory, and treatment* (pp. 415–451). New York: Springer-Verlag.
- Harris, K. R., Schmidt, T., & Graham, S. (1998). Every child can write: Strategies for composition and self-regulation in the writing process. In K. R. Harris, S. Graham, & D. Deshler (Eds.), *Advances in teaching and learning. Vol. 2: Teaching every child every day: Learning in diverse schools and classrooms* (pp. 131–167). Cambridge: Brookline Books.
- Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Mahwah, NJ: Erlbaum.
- Hayes, J. R. (2006). New directions in writing theory. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 28–40). New York: Guilford.
- Hayes, J. & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Erlbaum.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single subject research to identify evidence-based practices in special education. *Exceptional Children*, 71, 165–179.
- Intersegmental Committee of the Academic Senates (ICAS). (2002). *A statement of competencies of students entering California's public colleges and universities*. Sacramento, CA: ICAS.
- Lipsey, M. & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- MacArthur, C., Graham, S., & Fitzgerald, J. (2006). *Handbook of research on writing*. New York: Guilford.
- MacArthur, C., Schwartz, S., & Graham, S. (1991). Effects of a reciprocal peer revision strategy in special education classrooms. *Learning Disability Research and Practice*, 6, 201–210.
- McCormick, C. B. (2003). Metacognition and learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (Vol. 7, pp. 79–102).
- McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 115–130). New York: Guilford.
- Meltzer, L. (2007). *Executive function in education: From theory to practice*. New York: Guilford.
- National Commission on Writing (2003). *The neglected "R."* College Entrance Examination Board.
- National Commission on Writing (2007). *Writing and school reform*. College Entrance Examination Board.
- Nystrand, M. (2006). The social and historical context for writing research. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 11–27). New York: Guilford.
- Paris, S. G. & Lindauer, B. K. (1982). The development of cognitive skills during childhood. In B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 333–349). Englewood Cliffs, NJ: Prentice-Hall.
- Persky, H., Daane, M., & Jin, Y. (2003). National assessment of educational progress: 2002 report card for the nation and the states. Washington, DC: US Department of Education.
- Pressley, M. & Harris, K. R. (2006). Cognitive strategies instruction: From basic research to classroom instruction. In P. A. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 265–286). New York: Macmillan.

- Prior, P. (2006). A sociocultural theory of writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 54–66). New York: Guilford.
- Rogers, L. & Graham, S. (2007). A meta-analysis of single-subject design writing research. Manuscript in preparation.
- Rohman, G. (1965). Pre-writing: The stage of discovery in the writing process. *College Composition and Communication*, 16, 106–112.
- Sawyer, R., Graham, S., & Harris, K. R. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology*, 84, 340–352.
- Scardamalia, M., & Bereiter, C. (1986). Written composition. In M. Wittrock (Ed.), *Handbook of research on teaching*, 3rd ed. New York: Macmillan, pp. 778–803.
- Scruggs, T. & Mastriopieri, M. (2001). How to summarize single-participant research: ideas and applications. *Exceptionality*, 9, 227–244.
- Sitko, B. M. (1998). Knowing how to write: Metacognition in writing instruction. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 93–116). Hillsdale, NJ: Erlbaum.
- Swedlow, J. (1999). The power of writing. *National Geographic*, 196, 110–132.
- Torrance, M., Fidalgo, R., & Garcia, J. (2007). The teachability and effectiveness of cognitive self-regulation in sixth grade writers. *Learning and Instruction*, 17, 265–285.
- Tracy, B., Reid, R., & Graham, S. (2007). *Teaching young students strategies for planning and drafting stories: The impact of Self-Regulated Strategy Development*. Paper submitted for publication.
- Wong, B., Harris, K. R., Graham, S., & Butler, D. (2003). Cognitive strategies instruction research in learning disabilities. In L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of research on learning disabilities* (pp. 323–344). New York: Guilford.
- Wong, B., Hoskyn, M., Jai, D., Ellis, P., & Watson, K. (2007). *The comparative efficacy of two approaches to teaching sixth graders opinion essays*. Manuscript submitted for publication.
- Zimmerman, B. (1989). A social cognitive view of self-regulated learning. *Journal of Educational Psychology*, 81, 329–339.
- Zimmerman, B. & Reiserberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology*, 22, 73–101.

9 Writing is Applied Metacognition

*Douglas J. Hacker, Matt C. Keener, and
John C. Kircher*
University of Utah

Writing is the production of thought for oneself or others under the direction of one's goal-directed metacognitive monitoring and control, and the translation of that thought into an external symbolic representation.

Rather surprisingly, in the extensive research on writing, a definition of the writing process is rarely provided. Numerous investigations have been conducted concerning the various components of writing, the influences on writing, and what experts do that novices do not do when writing, but a definition of the writing process is seldom mentioned and often simply implied. Considering that a goal of science is to obtain understanding of the phenomena under investigation, the omission of an explicit definition in the literature is almost disturbing. Stephen Witte (1992) suggested that questions about the nature of writing are not usually discussed because no theory bridges gaps among textual, cognitive, and social dimensions of writing, and because of the uncertainties that enshroud writing as a process or product. Roy Harris (2000) admits that the “bare idea of writing” is not a popular topic “because there are so many vested interests concerned to promote ideas of writing that are far from bare” (p. 8). However, he goes on to say that the intellectual pursuit to reconstruct the “bare idea of writing” might be worth pursuing.

In this chapter, we have taken up that pursuit by starting with an explicit definition of the writing process. We divide that definition into parts and then proceed to describe and explain what those parts mean. Next, we argue that not only are metacognitive monitoring and control essential components of writing, but that writing is applied metacognition. From this argument, we propose a metacognitive theory of writing from which our definition of writing is derived. Finally, we end with a brief description of a new methodology for investigating writing, which has produced data that have contributed to our definition and metacognitive theory of writing.

Effective writing requires audience awareness. We are aware of different camps that investigate writing and the theoretical differences that separate those camps. Researchers in the cognitive camp, such as John Hayes (1996, 2006) or Deborah McCutchen (2006), when reading our definition may resonate with the word “metacognitive” and read with interest. Researchers in the sociocultural camp, such as Paul Prior (2006), who believe that “sociocultural theories represent the dominant paradigm for writing research today” (p. 54) will read the word “metacognitive,” see no mention of context, and might decide to read no further. Researchers from a semiotic camp, such as the late Stephen Witte, may see the “Witte” name cited numerous times and read with curiosity. Researchers from a social-interactive camp, such as Martin Nystrand (1989, 2006), may read “production of thought for oneself or others” and hesitate, questioning whether the “or” should be “with.” As Harris (2000) said, we all have our vested interests to promote ideas of writing,

and we admit that we have our own vested interests. But a goal in our pursuit of that “bare idea of writing” is to propose a new theory of writing that incorporates critical elements from all the various writing camps.

Thomas Kuhn (1970) identified two conditions that new paradigms must meet: The new paradigm must resolve some generally recognized problem that has not been resolved by earlier paradigms, and the new paradigm must preserve a relatively large part of the problem-solving ability of its predecessors. Although hardly a new paradigm, we believe that our theory of writing meets both conditions: Our theory resolves problems in existing writing theories by re-conceptualizing writing as primarily a metacognitive process in which the production of text is the result of a person’s goal-directed monitoring and control of their cognitive and affective states, and we preserve a great deal of the critical elements from other theories of writing.

A Definition

In this section, we divide our definition into four meaningful parts: (a) production of thought, (b) oneself or others, (c) goal-directed metacognitive monitoring and control, and (d) translation of thought into an external symbolic representation. We describe and explain what those parts mean with the goal of constructing the “bare idea of writing” in a way that bridges gaps among textual, cognitive, and social dimensions of writing and addresses whether writing should be defined as process or product. We end with our argument that writing is applied metacognition.

Production of Thought

Production of thought is the core of writing. Regardless of what symbolic form it takes, writing is imbued with thought. Whether it is scribble writing accompanied by a drawing, a grocery list, pictographs, an essay, a novel, or a scientific report, the goal of the author is to produce meaningful thought either by translating his or her thoughts into a tangible written form or by discovering meaning through this translation process (e.g., Flower & Hayes, 1984; Kellogg, 1994; Nystrand, 1989; Prior, 2006; Witte, 1992). Disagreements do arise as to whether the production of thought is a solitary cognitive act on the part of the writer or a joint social act on the part of the writer in negotiation with a reader. Those disagreements will be addressed more fully in the next section. The point being made here is only to argue for the first part of our definition of writing, which seems to be common to all modern theories of writing, “Writing is the production of thought . . .”

Oneself or Others

The issue of writing being a solitary cognitive process or a joint social process has been at issue, sometimes hotly contested, for at least 30 years. The dominance of the cognitive perspective of the 1970s gave way rather quickly to social perspectives with the works of Shaughnessy (1977), Nystrand (1982), Heath (1983), and Prior (1991), to name just a few (see Nystrand, 2006, for an excellent review of the social and historical context of writing research). More recently, some of the original proponents of the cognitive perspective on writing have acknowledged a more salient role for social influences (e.g., Flower 1994; Hayes, 1996).

On the one hand, one cannot deny that writing usually is a social artifact. The form and content of our writing is shaped by the conventions of our society and culture at a particular time in history (Bakhtin, 1981; Vygotsky, 1991). We write within a world bound by

those conventions, and even if an author breaks the writing conventions of a particular segment of society (e.g., James Joyce's stream of consciousness in *Ulysses*), the departure from conventions is still understood and judged by that society's established writing conventions. We use a socially shared language, a socially shared writing system, socially shared artifacts, mores, values, and history, and a socially shared culture. And, to write effectively to another, one must always keep in mind the intended generalized social audience. In all these ways, we can say that writing is social.

On the other hand, we cannot ignore that writing also is a cognitive activity that occurs within an individual. Writing reflects the unique phenomenology of an individual. According to Charles S. Peirce, the writer links an object (i.e., concept, entity, idea, feeling, or emotion) to a sign or series of signs (Moore, 1972). Linking an object to a sign requires that the writer first brings personal understanding to the object. What is the object that I hope to describe in words? What are its properties? What does it mean to me and others? What does it do? What does it not do? The personal understanding must then be constructed through a system of signs such that a link is established between the object and sign. Peirce contends that the sign must activate the object and the object must activate the sign (Moore, 1972). Kellogg (1994) refers to this cognitive process as translating one's personal symbols into consensual symbols. Further, the signs selected by the writer (i.e., the consensual symbols) must transmit to the reader a meaning that at least approximates the original object of the writer.

The approximation of the writer's thoughts by the reader requires a negotiation of the "intellectual and emotional space between the 'self' and the 'other,' between the individual and the social" (Witte, 1992). Therefore, the chances that the reader's configuration of meaning will match exactly that of the writer are fairly minimal. Witte goes on to say:

Hence what I write to mean will likely be read by someone else—in some way—to mean something else. The downside, of course, is that we—writer and reader—can only ever communicate approximately, which is generally the case between a "self" and an "other." (p. 288)

A similar sentiment was expressed about 2500 years earlier by Plato (1961) in the *Phaedrus*:

Socrates: You know, Phaedrus, that's the strange thing about writing, which makes it truly analogous to painting. The painter's products stand before us as though they were alive, but if you question them, they maintain a most majestic silence. It is the same with written words; they seem to talk to you as though they were intelligent, but if you ask them anything about what they say, from a desire to be instructed, they go on telling you just the same thing forever. And once a thing is put in writing, the composition, whatever it may be, drifts all over the place, getting into the hands not only of those who understand it, but equally of those who have no business with it.

(*Phaedrus*, 275e)

If chances are minimal that we can fully negotiate a meaning in our writing with another, and an "intellectual and emotional space" will likely always exist, there is some part of writing that is not social. That is, even the best of writers with the best of intentions to communicate meaning can produce words that are meaningful to him or her but will fail to be meaningful to another. This failure to communicate occurs, in part, because not all thought is dependent on language. We have spatial, visual, tactile, auditory, emotional, and other non-linguistic thoughts that do not translate easily or at all into a verbal form. In

these cases, not only must the writer bring interpretation into play to understand his or her own thoughts (verbal and non-verbal), a reader must bring interpretation into play to understand the writer's interpretation (Gadamer, 1993; Nystrand, 1982). Witte comes to the conclusion that the arguments between cognitive and social approaches "may ultimately point as much to the inadequacy of a purely 'social perspective' as it does to the inadequacy of the Flower and Hayes cognitive process theory" (p. 262). This seems to be a reasonable conclusion. To address the inadequacies in each theoretical perspective, a viable theory of writing perhaps must incorporate both perspectives.

Writing can be a solitary cognitive act of producing meaning for oneself. For example, Leonardo da Vinci's personalized system of writing was known only to him for the purposes of expressing only to his "self" his verbal and non-verbal thoughts. Or, writing can be a social act of producing meaning through negotiation with others. When writer and reader are face-to-face, the negotiation of meaning can be a powerful tool to enhance understanding for both writer and reader. But, when writer and reader are not in spatial, chronological, cultural, or historical contact, the negotiation of meaning occurs in more of an abstract sense, but it nonetheless carries with it a "potential for interaction" (Nystrand & Himley, 2001, p. 199). Thus, "Writing is the production of thought for oneself or others . . ."

Goal-Directed Metacognitive Monitoring and Control

Although there are declarative and procedural knowledge components to metacognition, we will focus mainly on the procedural components. These cognitive processes "monitor the selection and application as well as the effects of solution processes and regulate the stream of solution activity" (Kluwe, 1982, p. 204). Metacognitive monitoring is often described as an awareness of one's current thoughts or behavior, and metacognitive regulation or control is described as the modification of one's current thoughts or behavior. The importance of both components of metacognition has been recognized in the area of writing for at least 30 years. In the original Hayes and Flower (1980) model, the actions of three working memory processes, planning, translating, and reviewing, are controlled and coordinated by a monitor. Bruer (1993) stated that a well-developed monitor is necessary to keep the complex, recursive processes of writing manageable. Kellogg (1994) proposed that a writer must monitor and evaluate the progress of his or her thinking and writing using metacognition. Deliberate control of writing subskills was identified by Bracewell (1983) as important for the acquisition of new language subskills that are specific to writing and for solving problems that occur during writing. In an update of the Hayes and Flower model, Hayes (1996) maintained the importance of metacognitive control in writing, which is accomplished through a task schema that guides and controls the writing process; and Flower (1994) continued to acknowledge the importance of metacognition in the tacit or explicit control of cognitive acts that contribute to writing. Finally, Nystrand (1989) acknowledged that a monitor is necessary to coordinate the components of writing, but that "a true theory of writing requires more than an inventory of components; a theory of writing needs a principled explanation of the monitor itself" (p. 69). We agree with Nystrand on that point, and our goal, in part, is to provide that "principled explanation."

Every act of writing is an act of meaning production. Reading, re-reading, reflecting, and reviewing—processes traditionally associated with writing—serve as monitoring strategies to ensure that the production of meaning is in conformance with the author's goals for writing, and as long as the author's monitoring is accurate, this conformance is more likely to occur. Editing, drafting, idea generation, word production, translation, and revision serve as control strategies that are responsible for the actual production of meaning. In most cases of writing, by the time the first words have been written, the author has already

monitored and controlled the production of numerous ideas that are potentially appropriate to his or her goals for writing. Of these numerous ideas, only a select few have probably been translated into written text, although this varies with writing experience, with less experienced authors translating more of the ideas produced, and more experienced writers translating less. Bereiter and Scardamalia (1987) referred to the former as knowledge-telling writers and the latter as knowledge-transforming writers.

Online monitoring and control of writing continues uninterrupted until the author experiences a breakdown in meaning, which indicates that something in the text no longer conforms to his or her goals for writing, or additional meaning is simply not forthcoming. To re-establish meaning production, the author must exert control to ensure that any inappropriate mapping is remediated and further production continues. At a minimum this requires an orchestration of monitoring and control, involving a variety of strategies, such as accurately diagnosing the breakdown in meaning, reviewing what has been written, generating new ideas, and rewriting to produce new text that is in better conformance with the writer's purposes. Once the breakdown is resolved, the author again continues with writing until another breakdown in the production of meaning occurs. As larger units of text are completed, each larger unit can be monitored through a review process for meaning at a macro-level, and if the review indicates non-conformance to the author's goals, further control must be exerted. If the review indicates conformance to one's goals, meaning production continues until the author decides to stop.

Very likely, some portion of the monitoring and control of writing is tacit or implicit (Flower, 1994; Newell & Simon, 1973; Reder & Schunn, 1996). Within just the language production portion of writing, several monitoring functions occur, many requiring less than 100 ms (Postma, 2000). Moreover, control strategies such as planning, detecting, and diagnosing problems can be highly automated, implicit processes (Flower, 1994). The extent to which monitoring and control of writing is implicit is associated with the degree of writing expertise, with higher degrees of expertise associated with greater levels of implicit processing. For younger or inexperienced writers, language production may be the only implicit process available to them. This makes every other process involved in writing explicit, demanding of cognitive resources, and potentially an impediment to writing. However, it is important to keep in mind that although automatic processing can be a valuable asset to effective writing, explicit monitoring and control are essential components in the production of meaningful text. Only through explicit reading and re-reading, reviewing, revising, editing, and deliberate production and reproduction of text can the writer gain confidence that the written text is a good representation of his or her thoughts.

Granted that writing involves metacognitive monitoring and control of the production of meaning, we must ask the question, what meaning? We do not intend to delve into a philosophical discussion of the "meaning of meaning," but we do need to think about the meaning an author is trying to produce. The meaning produced during writing is directed by the writer's goals. The idea that human thought and behavior are best analyzed and understood in terms of people's goals has become a commonly accepted concept across disciplines (e.g., Cantor & Kihlstrom, 1987; Elliott & Dweck, 1988; Markus & Nurius, 1986). Carver and Scheier (1991) are perhaps two of the strongest proponents of this view: "In our view, virtually all of human behavior has this goal-oriented quality" (p. 168).

Goals are structured in a hierarchy, with higher-level goals being translated into lower-level goals (Conway, 2005). In the case of writing, an example of a higher-level goal would be to write a paragraph that describes how symbols differ from signs. The writer must then monitor and control a production of meaning that conforms to this goal. As the writer monitors and controls the production of text, he or she will adopt lower-level goals. For example, the writer may monitor for appropriate word choice, maintain coherent meaning

within and between sentences, or ensure proper grammatical structure. The knowledge necessary for successful goal-directed monitoring and control of meaning production is drawn from the writer's memory, but if the knowledge is insufficient or lacking, other sources must be sought (Carver & Scheier, 1991), such as other texts or people. The seeking of other sources appropriately illustrates the need identified by Witte (1992) for a theory of writing to bridge the textual, cognitive, and social dimensions of writing. The text emerges as a consequence of the knowledge of the writer and the social collaborations with others, all under the direction of the writer's goal-directed monitoring and control of the production of meaning.

Also important to note is that goals change as the task progresses; therefore, the focus of one's metacognitive monitoring and control changes as well. As a text grows, the task situation continually changes so that the writer is never thinking about the text in exactly the same way (Rijlaarsdam & van den Bergh, 1996). With each word written, the task environment presents a new stimulus to which the writer must generate a new response. This is not a passive behavioral relationship between an external environmental prompt and an associated response that is reinforced. This is an active within-person relationship, with the writer knowingly generating a stimulus (i.e., the written text), which is representative of his or her thoughts, and knowingly responding to that stimulus with further thoughts. As the writer's thoughts change, different goals are evoked (Carver & Scheier, 1991), and as the goals change so does the focus of the writer's metacognitive monitoring and control. Thus, the writer's goals set the stage for production of meaning and for the character and quality of the monitoring and control of that production. Therefore, "Writing is the production of thought for oneself or others under the direction of one's goal-directed metacognitive monitoring and control . . ."

Translation of Thought into an External Symbolic Representation

Writing is the author's production of meaning, and that production must be translated into an external symbolic representation. But what symbolic representation? We most often think of traditional, conventionalized writing systems; however, do these conventionalized writing systems adequately represent the meaning-making that humans are capable of producing? As mentioned previously, not all thought is dependent on language. We have spatial, visual, tactile, auditory, emotional, and other non-linguistic thoughts that do not translate easily or at all into a verbal form. Therefore, we need to consider the instantiation of meaning in text in ways that do not privilege traditional, conventionalized forms of language. Witte (1992) argues that if we restrict ourselves only to traditional language we will not be able to account for many different kinds of "written text," written text that is culturally determined and able to account for the "meaning-constructive and social-constructive dimensions of writing" (p. 249).

If we agree that writing is the production of meaning, then we must be open to other external symbolic representations of that meaning. Children's emerging writing, which is often a cross between pictures and conventionalized letters or letter strings, is (a) a translation of meaning that is certainly not entirely conventional but is nonetheless culturally-constructive as determined by the type and content of the pictures and letters drawn, (b) is meaning-constructive as determined by the ideas the child wishes to convey, and (c) is social-constructive as determined by the intended message that is sent to another. Guidebooks to birds make use of conventionalized writing systems, but without the pictures of each bird, the guidebooks would be useless. The usefulness of an operator's manual on how to fix a mechanical device depends as much on having the actual mechanical device to make sense of the manual as it does to have the manual to make sense of the mechanical

device (Kellogg, 1994). The manual/mechanical devices are cultural contrivances that are bound by meaning- and social-constructive processes. Finally, text messaging is a production of meaning that uses bits and pieces of a conventionalized writing system but goes beyond that system in a way that is bound to a specific culture, context, and is both meaning- and social-constructive. Consider the following example: “omg its ez 2b bffs so y ru not going 2 da part-t l8tr 2 nite ppl r saying ura baby 4 not going so u betr go K ttyl.* Although bits of conventionalized writing can be identified, the translation of the author’s meaning and the conveyance of that meaning are highly dependent on the social and cultural milieu in which it was written.

The translation of meaning in an external symbolic representation entails more than a simple choice of what writing system to use. If writing is the production of meaning, and meaning can be more broadly defined than linguistic meanings, then we need to think about a more broadly defined semiotic system (i.e., theory of signs and symbols) than the traditional, conventionalized systems that are commonly used (Witte, 1992). We must acknowledge the fact that with any narrowly defined semiotic, such as a conventionalized writing system, it is unlikely that we can perfectly represent the world through language, and it is also unlikely that we can communicate perfectly our personal representation to another (Sapir, 1961; Witte, 1992). But we should be willing to accept these shortcomings if we accept writing as both a conveyance of meaning *and* a discovery of meaning (Witte, 1992). The discovery occurs for both the writer and reader: For the writer, when he or she searches for just the right symbol or sign to represent his or her ideas; for the reader, when he or she attempts to translate that sign or symbol. Whether conveyance or discovery, “Writing is the production of thought for oneself or others under the direction of one’s goal-directed metacognitive monitoring and control, and the translation of that thought into an external symbolic representation.”

Writing is Applied Metacognition

In the foregoing discussion, we argued that metacognitive monitoring and control are essential components of writing. We take this argument one step further by proposing that writing *is* applied metacognitive monitoring and control. We will pursue this argument by using Stephen Witte’s (1992) question of whether writing should be defined as process or product. The distinction between process and product makes sense when we are readers of other authors; however, when we are readers of our own writing, the distinction between process and product is not so clear. In this case, the process of writing is a reflection of our thinking, and the product of writing is a reflection of our thinking. How can we look at our own writing and not also look at our own thoughts?

We will start our argument by drawing on the multitude of definitions of writing provided in the Oxford English Dictionary (2008). There are a variety of definitions of writing as a noun (i.e., the product of writing) and as a verb (i.e., the process of writing). As a noun, we have selected . . .

The using of written characters for purposes of record, transmission of ideas, etc.
Expression of thoughts or ideas in written words; literary composition or production.
Words, letters, etc., embodied in written (or typewritten) form.

* We would like to thank Emily Hacker for sharing her text messaging with us. The message reads: “Oh my gosh, it’s easy to be best friends so why are you not going to the party later tonight? People are saying you are a baby for not going, so you better go, OK? Talk to you later.”

As a verb . . .

- To give expression to (one's feelings, thoughts, etc.) by means of writing.
- To express or present (words, etc.) in written form.
- To compose and set down on paper (a literary composition, narrative, verse, etc.).

As readers of another person's writing there is a clear distinction between product and process. The product lies before us. We can access the author's written letters, characters, and words, read them, and create meaning from them. However, the process of writing, during which the author translated his or her thoughts or ideas in written form, is hidden in the product. We can infer the process from the product, but with any inference made during reading, it is difficult to know how much our own knowledge and beliefs are influencing the intended text. As readers, we must reconstruct how the author used the processes of writing to express the meaning that he or she had in mind, and we must do this in a way that resembles as closely as possible the author's meaning and purpose for writing (i.e., engage in hermeneutical interpretation). Indeed, the goal of literary criticism is to expose the writer's thinking through an analysis of the writer's finished text (Kellogg, 1994). Sometimes our interpretations may not be accurate representations of the author's process of writing, and as was mentioned previously in the quotes from Witte and Plato, what we "write to mean will likely be read by someone else—in some way—to mean something else." *Good enough* interpretations may be the best we can hope for, although *good enough* at times may be acceptable and even desired.

As readers of our own writing, the distinction between product and process becomes blurred. Writers are in a privileged position. Not only do they generate the thoughts that they wish to write, and monitor and control that generation of thoughts, but they translate those thoughts into writing, and they monitor and control that translation. The only difference between monitoring and controlling the thoughts being generated and the monitoring and controlling of the thoughts being written is that the thoughts being written likely represent only a subset of the thoughts being generated. Not all thoughts make it to print, but as writers we know what thoughts do and what thoughts do not. We have no need to infer the intended meaning. We know the meaning, and even if that is in question, we are privy to the interpretations that may be used to find that meaning. Reading, re-reading, reflecting, and reviewing are used as monitoring strategies of our own thoughts. Editing, drafting, idea generation, word production, translation, diagnosing, and revision are used as control strategies of our own thoughts. The monitoring and control of our own thinking is metacognition. Writing is applied metacognition. Joan Didion (1980) put it much more eloquently, "I write entirely to find out what I'm thinking, what I'm looking at, what I see and what it means" (p. 335).

A Metacognitive Theory of Writing

The Nelson and Narens (1990) model of metacognition has served as a versatile theoretical framework for the conceptualization of metacognition and as a heuristic for further theorizing and empirical research. Three principles underlie their model: (a) Mental processes are split into two or more specifically inter-related levels, a cognitive and a metacognitive level; (b) the metacognitive level contains a dynamic model of the cognitive level; and (c) there are two dominance relations called control and monitoring, which are defined in terms of the direction of flow of information between the meta-level and the object-level. Figure 9.1 illustrates a modification of the Nelson and Narens model as applied to writing. Based on the state of the dynamic model of the object-level at the

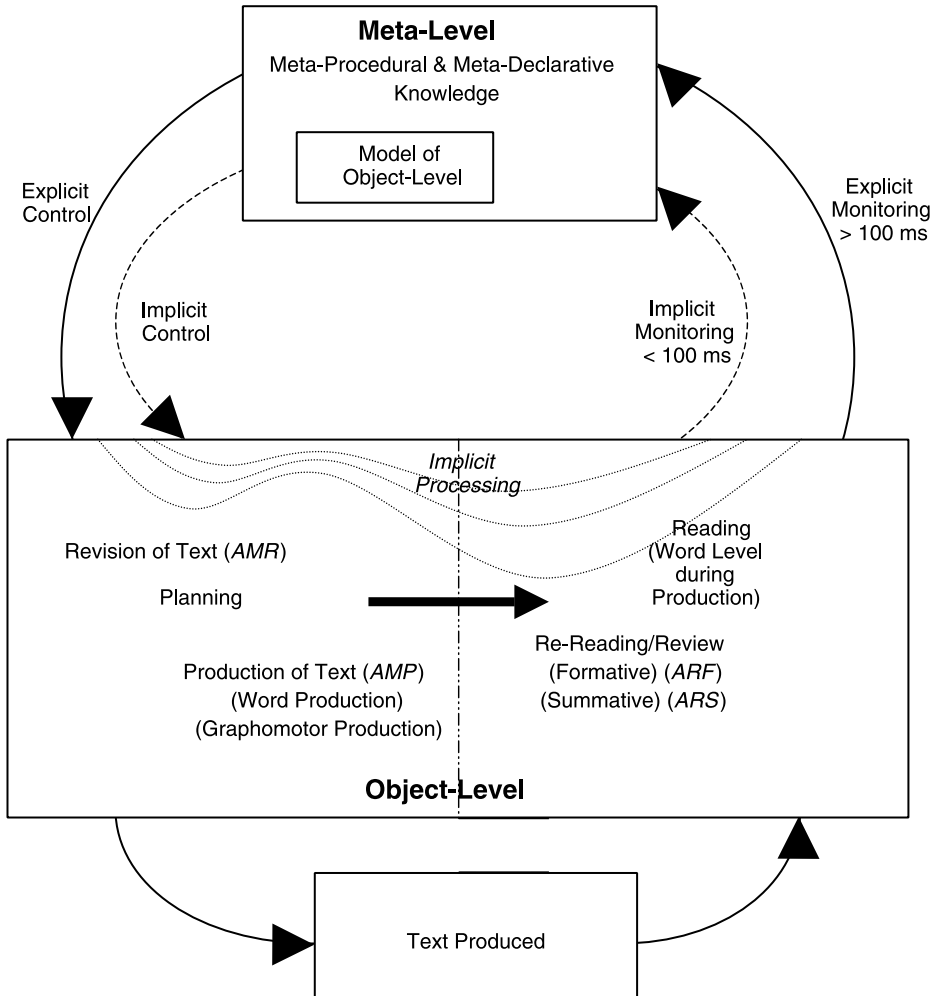


Figure 9.1 Metacognitive model of writing.

meta-level, control information flows from the meta-level and modifies the state of the object-level process or knowledge or changes the object-level altogether. Modifications or changes at the object-level are monitored at the meta-level so that modifications to the model of the object-level can be made. A change in state of the model of the object-level can lead to additional control information flowing to the object-level.

As was mentioned previously, some portion of the monitoring and controlling of writing is tacit or implicit (Flower, 1994; Newell & Simon, 1973; Reder & Schunn, 1996). For example, detection and repair of an error at the phonemic level can occur about 100 ms before there is overt realization of the error (Postma, 2000). Moreover, highly practiced strategies, such as planning, diagnosing, or reviewing can be automatized (Flower, 1994). Therefore, the monitoring and control of cognitive processes and knowledge at the object level can be either explicit or implicit. This is illustrated in Figure 9.1 by the solid and dashed lines indicating the explicit and implicit flow of information in monitoring and control. Deliberate and intentional writing relies more on explicit than implicit processing, although writing processes that are put into action at the object-level might always include

some implicit processing. This is illustrated in Figure 9.1 by the dashed curved lines within the object-level.

Writing is an act of meaning production, and during the goal-directed monitoring and control of meaning production, the author uses reading, re-reading, and reviewing as monitoring strategies to ensure that the production of meaning is in conformance with the author's goals for writing. Word production, revision of text, planning, and production of text serve as control strategies that are responsible for the actual production of meaning. Ideas produced as a consequence of the control strategies are monitored and compared against the goals the author has established at the meta-level. Some of the ideas are written and end up in the text but others do not. The author must continually monitor not only the ideas produced but the selected ideas that end up in the text so that intended goals are being met.

Hayes and Flower (1980) originally posited that writing processes are hierarchical in nature. Three working memory processes—planning, translating, and reviewing—work in concert during writing, and within each of these three processes are embedded sub-processes: Planning includes the generation and organization of content and goal setting; reviewing includes reading and editing. Furthermore, each of these mental acts may occur at any time in the writing process and are considered recursive (Hayes & Flower, 1980).

We agree with the idea of recursion, but we contend that rather than a hierarchy of writing processes as proposed by Hayes and Flower, which suggests that processes higher in the hierarchy have precedence over processes lower in the hierarchy, the writer either implicitly or explicitly engages whatever control or monitoring process is needed to maintain meaning production at the moment of writing. Moreover, although control processes are distinct from monitoring processes, the line that separates the two is often tenuous. This distinction is represented in Figure 9.1 by the dashed line between control and monitoring at the object-level. Consider these examples. As writers plan, they simultaneously monitor their plans for meaning and conformance to their goals, which can result in immediate alteration of the plans, further review of their plans, or writing of a few words or entire strings of words. As writers translate their thoughts into words, they simultaneously monitor those words for meaning and conformance to their goals, which can lead to further translation, major or minor revision of what was written, or regressing back to earlier sections for review and possible revision. As writers re-read what they have written, the re-reading can serve as a stimulus for the production of more words. A reorganization of a sentence for better meaning can lead imperceptibly to re-reading the prior sentence, which can lead imperceptibly into altering goals for writing. Using either pen and paper, keyboard and monitor, or crayon and the bedroom wall, the writer's goal is to produce meaning in an external symbolic form, and whatever process will help in that production will be used whenever it is needed. Each monitoring or control process in use contains within it the potential for every other process, and what determines the selection of process is whether the writer's intended meaning is being produced.

Moreover, as the text develops the task situation continually changes. The writer never attends to the same meaning in the same way anytime during writing (Rijlaarsdam & van den Bergh, 1996). Therefore, the quality and character of any writing process, whether it involves control or monitoring of meaning production, could change considerably depending on how the production of meaning proceeds for the writer. Reviewing occurring at one time may be much different from reviewing that occurs at another time. Revision of the text at the earlier stages of writing will be much different from revision that occurs at later stages (Rijlaarsdam & van den Bergh, 1996). And this can be said of every writing process. As was mentioned earlier, each monitoring or control process contains within it the

potential for every other process, and this potential changes as the writing task situation changes.

The complexities and difficulties of writing stem in large part because of what seems to be a completely chaotic process. Writing may appear to be much like the “Blooming, buzzing, confusion” that Williams James (1890) described of our world without the benefit of perceptual categories. But humans are meaning-making creatures, and as long as the purpose of writing is to produce meaning, the blooming, buzzing, confusion of writing will result in meaning for the author and hopefully for the reader.

New Methodology to Investigate Writing

With the complexities presented by even the simplest writing task, the investigation of writing is a difficult task. Hayes and Flower (1980) introduced the think-aloud methodology to writing, which had been used extensively to gain access to people’s cognitions as they problem solve (e.g., Bereiter & Scardamalia, 1987; Chi, Lewis, Reimann, & Glaser, 1989; Montgomery & Svenson, 1989; Pressley & Afflerbach, 1995). Hayes and Flower’s research was seminal and led to innumerable studies of writing.

Ronald Kellogg (1987) provided critical understanding of writing by investigating processing time and effort people invested in the various cognitive processes that had been identified by Hayes and Flower. He made use of the directed retrospection technique (Ericsson & Simon, 1980) in which writers were interrupted at various times during writing and asked to retrospectively identify the writing process in which they were engaged prior to the interruption. The retrospective reports indicated what processes were involved in writing, and they allowed investigators to calculate the proportion of time devoted to each process. Cognitive effort was assessed by secondary response times (Kahneman, 1973). This was measured by recording the amount of time writers required to identify the writing process in which they were engaged. Cognitive effort was defined in terms of the increase over a baseline response time calculated when writers were not engaged in writing.

Although these measures have increased our understanding of writing, they do not entirely reveal the moment-to-moment processes involved while writing. Problems with retrospective reports and secondary response times include their coarse granularity, lack of specificity, unnatural writing experiences, invasiveness, and disjointedness in providing a cohesive representation of writing. As an alternative to these approaches, we have developed a new methodology to investigate the online writing behaviors that are essential for understanding the role that metacognitive monitoring and control play in writing. Our new methodology makes use of eye-tracking technology, which, although not problem-free, is considered the state of the art for online investigations of language comprehension (Boland, 2004; Carreiras & Clifton, 2004; McConkie, Hogaboam, Wolverton, Zola, & Lucas, 1979; Rayner, 1998). Eye tracking provides continuous measures of processing time, attention, and effort, does not disrupt the reader or writer from the main task, produces data reflecting attentional shifts in periods of time as short as a few ms, can pinpoint comprehension problems at a word and even intra-word level, and provides a more natural way of examining word and sentence processing (Boland, 2004; Carreiras & Clifton, 2004).

Traktext

We developed an extension of eye-tracking technology to track eye movements during writing. Our computer program, called Traktext, displays several windows on a computer

monitor: some windows present static texts used in traditional eye-tracking analyses of reading behaviors, and one presents a word processing program used to record participants' writing behaviors. We continuously record participants' eye movements as they freely navigate among the reading and writing windows. The dynamic nature of writing is captured offline at the word level by generating a bitmap of the writing window for each word that the participant types, edits, or deletes. The derived spatial coordinates for each word are used in conjunction with the eye positions recorded 60 times per second by the eyetracker to measure the number and duration of eye fixations on the word.

Our Traktext program provides numerous measures of the writing process. A record of each and every word written, deleted, or edited is compiled along with the amount of time spent writing each word (measured in ms/character) and the amount of time spent reading and re-reading each word (also measured in ms/character). Cognitive effort allocated to the writing and reading of each word is measured using pupil diameter. Numerous studies have shown that the greater the demand on cognitive capacity, the more the pupil dilates (e.g., Ahern & Beatty, 1979; Granholm, Asarnow, Sarkin, & Dykes, 1996). Kahneman (1973) has shown that pupillometric response is a reliable and sensitive psychophysiological index of the momentary processing load during performance of a wide variety of cognitive activities, including attention, memory, language, and complex reasoning. Moreover, pupillometric response shows within-task, between-task, and between-individual variations in processing load and cognitive resource capacity (Kahneman, 1973).

In our current research paradigm, we are interested in both reading and writing behaviors during problem solving; however, for the purposes of this chapter, we will focus exclusively on writing behaviors. Traktext first presents participants with a simple pre-problem exercise that involves both reading and writing. Participants read a short prompt about the Jack and Jill nursery rhyme and then write as much as they can remember about the nursery rhyme. This allows us to obtain baseline rates for both keyboarding speed and pupil diameter for reading, language production, and keyboarding behaviors. These baseline rates are used as a standard from which writing and reviewing measures on the actual problem are calculated.

Following the baseline task, participants are presented with a text that describes a problem about gravity. It is a classic problem often used in refutational-text research about two balls falling from the same height but with different trajectories (Guzzetti, Snyder, Glass, & Gamas, 1993). Participants are asked which of the two balls would hit the ground first and why. After writing their answers, participants are presented with another text that contains information that can be used to help solve the problem. Once finished with the informational text, participants can then revise their answers in any way they wish and navigate freely between the text and the problem as they revise. Participants are allowed as much time as they wish to complete the task.

Analysis of Writing

Using these methods to analyze writing, we have identified six distinct processes: (a) area of monitoring and revision (AMR); (b) area of monitoring and production (AMP), which includes word (language) production and graphomotor production; (c) planning, which includes generating ideas, organizing, and goal setting (Hayes & Flower, 1980); (d) area of review—formative (ARF); (e) area of review—summative (ARS); and (f) reading at the point of production. Each of these is summarized in Table 9.1 and illustrated as a monitoring or control process in Figure 9.1.

Table 9.1 Control and Monitoring Processes, Acronyms, and Definitions

<i>Control strategies</i>	<i>Acronym</i>	<i>Definition</i>
Area of monitoring and production	AMP	Contiguous area of text produced prior to a writer making a saccade beyond 12 characters or spaces from his or her last fixation on a currently produced word to another part of the text or prior to a navigation off the text
Planning		Generating ideas, organizing ideas, goal setting, and re-setting goals
Area of monitoring and revision	AMR	Area of the text in which a writer makes a revision
<i>Monitoring strategies</i>	<i>Acronym</i>	<i>Definition</i>
Area of review—formative	ARF	Area of text in which the writer re-reads beyond 12 characters or spaces from the last point of production and then continues with writing
Area of review—summative	ARS	Area of text re-read just before the writer finishes with writing or before additional information is encountered that could lead to a change in the answer
Reading at point of production		Reading the current word that is written and one or two words prior to the current word

An AMR is an area of the text in which a revision occurs. Revision is broadly defined according to Fitzgerald (1987, p. 484):

Revision means making any changes at any point in the writing process. It involves identifying discrepancies between intended and instantiated text, deciding what could or should be changed in the text and how to make the desired changes, and . . . making the desired changes. Changes may or may not affect the meaning of the text, and they may be major or minor.

From this definition and from our description of our metacognitive model of writing, it is clear that revision involves both monitoring (“identifying discrepancies between intended and instantiated text”) and control processes (“making the desired changes”). The line that separates control from monitoring is not always distinct. As we suggested earlier, each monitoring or control process contains within it the potential for every other process, and what determines the selection of process is whether the writer’s intentions for meaning are being met.

We define an AMP as a contiguous area of text produced prior to a writer making a saccade beyond 12 characters or spaces from his or her last fixation on a currently produced word to another part of the text or prior to a navigation off the text. We selected 12 characters or spaces because this distance has been typically defined as the amount of text that can be in foveal view during a fixation (Rayner, 1998). Thus, moving a gaze beyond 12 characters or spaces indicates that the writer has made a saccade and likely is focusing his or her attention on parts of the text that are not currently being produced.

Figure 9.2 illustrates a participant’s writing that followed immediately after her reading of the problem on gravity. The ovals on each line indicate eye fixations. Lines a, b, and c constitute the ending of the writer’s first AMP. The writer has continuously produced the words “I think that the juggler who was standing still would have the” and in Line a has

a)	I think that the juggler who was standing still would have the fastest	AMP1 19,467 ms -2,992 ms 1.09
b)	I think that the juggler who was standing still would have the fastest ball .	
c)	I think that the juggler who was standing still would have the fastest ball. The	
d)	I think that the juggler who was standing still would have the fastest ball. THE juggler	ARF1 1,417 ms 117 ms 1.11
e)	I think that the juggler who was standing still would have the fastest ball. The juggler who	
f)	I think that the juggler who was standing still would have the fastest ball. The juggler was who was	AMP2 22,150 ms -6,457 ms 1.07
g)	I think that the juggler who was standing still would have the fastest ball. The juggler who was speeding	
h)		
i)		
j)		
k)		
l)	I think that the juggler who was standing still would have the fastest ball. The juggler who was speeding around the ring has some added force that I don't really understand,	
m)	I think that the juggler who was standing still would have the fastest ball. The juggler who was speeding around the ring has some added force that I don't really understand,	ARF2 1,250 ms 834 ms 1.07
na)		
nb)		
nc)		
nd)		
ne)		
nf)		
ng)		
nh)		
ni)		
nj)		
nk)		
nl)		
nm)		
nn)		
no)		
np)		
aq)	I think that the juggler who was standing still would have the fastest ball. The juggler who was speeding around the ring has some added force that I don't really understand, happening with his ball, not just gravity .	AMP3 8,917 ms -3,914 ms 1.03
ar)		
as)		
at)		
au)		
av)		
aw)		
ax)		
ay)		
az)		
ba)	I think that the juggler who was standing still would have the fastest ball. The juggler who was speeding around the ring has some added force that I don't really understand, happening with his ball, not just gravity.	ARS1 13,533 ms 8,733 ms .94

Figure 9.2 Writing sample illustrating eye fixations, writing processes, and descriptive statistics on processes.

produced “fastest.” As indicated by the fixations, the writer is currently fixating on the word being produced, “fastest,” and on the previous word “the.” In Line b, the writer is fixated on the word being produced, “ball,” and on the previous word, “fastest.” Fixating on the word being produced and one or two words prior to that word is commonly observed in our participants. Writers often appear to read the last word that is written as a lead-in to the word that is about to be produced. This is represented in our model of writing in Figure 9.1 as Reading (Word Level during Production). Looking at the word previous to the word produced may serve as a micro-review process that the writer uses to maintain the stream of thought contributing to continued production of meaning. In Line c, the word “The” is produced and fixated, which ends the AMP because the writer then regresses in the AMP, and as is shown in Line d, begins an ARE, which is defined as a review of text that lies beyond 12 characters or spaces from the last point of production. In other words, the writer made a saccade from the point of production to another portion of the text. This is a formative area of review because the review leads to further production of their current answer. In contrast, an ARS is the final review of text before finishing the

task or encountering additional information that could change their current understanding and answer.

The values presented with the AMP1 in Figure 9.2 indicate that the production of AMP1 took 19,467 ms, which is 2,992 ms faster than the writer's writing base rate, and the 1.09 is a ratio of the average pupil diameter during the AMP as compared to the baseline pupil diameter that was measured during the pre-problem exercise. The 1.09 indicates that greater cognitive effort was being expended as compared to the writer's baseline value. As an approximate estimate of what a 9% increase in pupil diameter means, Hess and Polt (1964) conducted pupil diameter studies in which they produced a rough calibration of pupil increase and cognitive effort. A 5% increase in pupil diameter is approximately the amount of increase when a person begins to solve a fairly easy multiplication problem such as 7×8 . A 10% increase is approximately the amount of increase when a person is solving a problem such as 8×13 . A 20% increase is approximately the amount of increase when solving a difficult problem such as 16×23 . Therefore, the writer produced AMP1 relatively quickly and showed a small increase in cognitive effort as compared to writing the Jack and Jill rhyme. A reasonable interpretation of AMP1 is that the writer likely had mentally produced his or her answer while reading the problem on gravity and quickly wrote her answer. This would be an illustration of a writing style identified by Bereiter and Scardamalia (1987) as knowledge telling, that is, the writer already has the knowledge formed and needs only to write it.

The values presented by the ARF1 indicate that ARF1 took 1,417 ms, which is 117 ms slower than the writer's reviewing base rate, and that the ratio of the writer's pupil diameter was 1.11. Thus, the writer interrupted the AMP to make a quick review of what she had written, and because of the increase in pupil diameter, was using a moderate amount of cognitive effort to conduct the review. The ARF is then interrupted to continue a second AMP in which "who," "was," "speeding," etc. were produced. In Line s, the second AMP is fully represented. Similar to AMP1, the values indicate that the production of AMP2 was fairly quick, taking 22,150 ms, which is 6,457 ms faster than the writer's base rate, and that the ratio of the writer's pupil diameter was 1.07, which indicates that less cognitive effort was being expended as compared to AMP1 and ARF1, but more than what was required for the Jack and Jill nursery rhyme. These values further support our interpretation that the writer was engaging in knowledge-telling writing in that she likely had already mentally produced her answer while reading the problem and quickly wrote her answer.

We are able to change our unit of analysis from the process level to the word level, and it is interesting to note that as the participant was writing the words "I don't really understand" in Line s, her pupil diameter ratios for the words were 1.11, 1.08, 1.12, and 1.11, respectively. These ratios indicate that she was expending a moderately high level of cognitive effort as she was attempting to understand her quandary in solving the problem.

After writing the word "understand," the writer interrupts her production for another short ARF that lasts 1,250 ms, which is 834 ms slower than the writer's base rate, and continues to expend a moderate amount of cognitive effort as indicated by the 1.07 pupil diameter ratio. ARF2 is followed immediately with another AMP in which "happening with his ball, not just gravity" is written. AMP3 is written quickly, taking only 8,917 ms, which is 3,914 ms faster than the base rate, and the pupil diameter ratio decreases to 1.03, indicating a slight decrease in cognitive effort. Possibly, during ARF2, the writer came to realize that her lack of understanding of the problem was due to some component other than gravity, and then used a portion of the ARF (i.e., the additional 834 ms) to plan her thoughts that were expressed in AMP3.

Finally, she ends her answer with an ARS. She reviews selected portions of her text one

last time, and even though she spends 13,533 ms reviewing, which is 8,733 ms more than her base rate, her pupil diameter ratio is quite low, 0.94, which indicates that there is a sharp decrease in her cognitive effort at this point. ARS1 ends with her decision to read the informational text on gravity, so perhaps the writer resigned herself to the idea that she was unable to solve the problem and would end her answer and consult with the informational text for further input. She then reads the informational text, returns to her writing, and through a series of AMRs, AMPs, ARFs, and ARSs, completely changes her writing to reflect a correct answer.

With this particular research paradigm, it is difficult to determine how much time was spent planning, because at least a portion of her planning was done during the reading of the task and the informational texts. However, within the AMPs, ARFs, ARSs, and AMRs, 14 seconds were unaccounted for, which could have been used for planning. In the future, we intend to use a paradigm that will more clearly isolate planning time.

Our methodology offers advantages over the traditional methodologies of think-aloud (Hayes & Flower, 1980) and directed retrospection (Kellogg, 1987). By investigating writing online, we get a more naturalistic representation of the processes in which writers engage as compared to think-aloud or retrospective report paradigms that may have unintended effects on the writing process. Sometimes we tell more and sometimes less than what we know (Nisbett & Wilson, 1977). As compared to self-report, eye-tracking methodology provides measures of behavior and physiology. Such measures may offer new insights into the kinds of processes that underlie the production of text, the transitions between processes, the timing of those processes, and the effort required to engage and maintain those processes.

Earlier, we had discussed how every writing process contains within it the potential for every other process and that changes can occur in a seemingly chaotic fashion. In our illustration, as well as with other participants we have observed writing, the writer went from producing text, jumped to reviewing, went back to producing text, jumped again to reviewing, back to producing, and finally back to reviewing. All of this occurred in 66.7 seconds and resulted in the production of 38 words. Some of these processes occurred in 1.25 seconds, and some represent the lion's share of time, such as production of text. Hayes and Flower (1980) discussed the recursiveness of writing, and our illustration shows that not only is writing recursive, it is furiously recursive at times. Moreover, we found evidence for three types of reviewing: a micro-review that occurs at the point of word production, a review that pops up seemingly unpredictably during word production, and a review that occurs at the end of word production. Likely, each review serves a different purpose, and our future research will investigate these purposes. Finally, measures of processing time and cognitive effort fluctuate considerably during writing. Both of these measures have been used in prior research to make inferences about the nature of the processes writers use. In our illustration, we identified areas of writing that went smoothly, efficiently, and seemingly without taxing the writer—areas of writing that would fit with the knowledge-telling concept of Bereiter and Scardamalia (1987). We also identified areas of writing that seemingly taxed the writer and led the writer to seek additional resources—areas that would fit with the knowledge-transforming concept of Bereiter and Scardamalia. We are only at the beginning of refining this writing methodology and disseminating it. The potential for further investigations with this technology are considerable.

Conclusion

In this chapter, we have provided a definition of writing that begins to bridge the gaps among textual, cognitive, and social dimensions of writing (Witte, 1992). Our definition

derives from our proposed theory of writing that conceptualizes writing as a metacognitive process in which meaning is produced under the direction of metacognitive monitoring and control processes. In turn, our writing theory was influenced by the new ways in which we are investigating writing. Writing requires both thinking and thinking about that thinking, and the best way to derive a better understanding of the complexities of writing is by getting as proximal to thinking as possible. Think-aloud protocols and directed retrospection have done well at this, but investigating writing at the moment of production is getting us perhaps as close to the thinking and the thinking about thinking as is possible. As in any science, observations lead to the development of theory, and theory leads to definitions. Our online observations of writing have led to changes in existing theories of writing, and the changes have led to a revised definition of writing. Our overall goal has been to get closer to that “bare idea of writing.”

Writing can be a solitary cognitive act of producing meaning for oneself, and writing can be a social act of producing meaning through negotiation with others. The very symbols that are used to express ideas, the manner in which the symbols are arranged, and the ways those symbols are interpreted by the writer and reader are socially, culturally, and historically bound. These aspects of writing cannot be ignored. But we also cannot ignore that there is a mind/brain that stores, manipulates, and uses the symbols for oneself or makes them available for others to use.

Kuhn (1970) proposed that new paradigms must meet two conditions: The new paradigm must resolve some generally recognized problem that has not been resolved by earlier paradigms, and the new paradigm must preserve a relatively large part of the problem-solving ability of its predecessors. Several writing theories that are still current consist mostly of lists of component processes that contribute to writing, such as generating ideas, organizing ideas, planning, reviewing, reading, content and rhetorical problem solving. These processes are still a part of our theory—although we conceptualize them as writing strategies—and therefore, a large part of the problem-solving ability of prior theories is maintained.

The problem that remains from earlier paradigms is that writing theories have not convincingly described how these component processes are coordinated under the direction of a monitor. “A true theory of writing requires more than an inventory of components; a theory of writing needs a principled explanation of the monitor itself” (Nystrand, 1989, p. 69). We have addressed this problem by providing a principled explanation of the monitor by re-conceptualizing writing as primarily applied metacognition in which the production of text is the production of meaning that results from a person’s goal-directed monitoring and control of their cognitive and affective states. Moreover, each monitoring or control process contains within it the potential for every other process, and what determines the selection of process is whether the writer’s intended meaning is being produced. Online monitoring of writing behavior reveals rapid and erratic changes from one writing process to another, with variable time courses and fluctuations in cognitive effort. These observations are consistent with the idea that one writing process can lead directly to any other, and all result as a consequence of the writer’s goal-directed monitoring and control.

References

- Ahern, S. & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, 205, 1289–1292.
- Bakhtin, M. M. (1981). Discourse in the novel. *The dialogic imagination: Four essays by M. M.*

- Babtin (pp. 259–422) (M. Holquist, Ed. C. Emerson & M. Holquist, Trans.). Austin: University of Texas Press.
- Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 51–76).
- Bracewell, R. J. (1983). Investigating the control of writing skills. In P. Mosenthal, L. Tamor, & S. A. Walmsley (Eds.), *Research on writing: Principles and methods*. Albany, NY: State University of New York.
- Bruer, J. T. (1993). *Schools for thought*. Cambridge, MA: The MIT Press.
- Cantor, N. & Kihlstrom, J. F. (1987). *Personality and social intelligence*. Englewood Cliffs, NJ: Prentice-Hall.
- Carreiras, M. & Clifton, C., Jr. (2004). On the on-line study of language comprehension. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 1–14).
- Carver, C. S. & Scheier, M. F. (1991). Self-regulation and the self. In J. Strauss & G. R. Goethals (Eds.), *The Self: Interdisciplinary approaches* (pp. 168–207). New York: Springer-Verlag.
- Chi, M. T. H., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How student study and use examples in learning to solve problems. *Cognitive Science*, 13, 145–182.
- Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, 53, 594–628.
- Didion, J. (1980). Why I write. In W. Smart (Ed.), *Eight Modern Essays*. New York: St. Martin Press.
- Elliott, E. S. & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54, 5–12.
- Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87, 215–251.
- Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research*, 57, 481–506.
- Flower, L. (1994). Metacognition: A strategic response to thinking. In L. Flower, *The construction of negotiated meaning* (pp. 223–262). Southern Illinois University Press.
- Flower, L. S. & Hayes, J. R. (1984). Images, plans, and prose: The representation of meaning in writing. *Written Communication*, 1, 120–160.
- Gadamer, H. (1993). *Truth and method*, second revised ed. New York: Continuum.
- Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, 33, 457–461.
- Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly*, 28, 117–159.
- Harris, R. (2000). *Rethinking writing*. Bloomington, IN: Indiana University Press.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theory, methods, individual differences, and applications* (pp. 1–27). Mahwah, NJ: Lawrence Erlbaum Associate.
- Hayes, J. R. (2006). New directions in writing theory. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 28–40). New York: Guilford.
- Hayes, J. R. & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (p. 3–30). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. New York: Cambridge University Press.
- Hess, E. H. & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science*, 40, 1190–1192.
- James, W. (1890). *The principles of psychology*. New York: H. Holt.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory & Cognition*, 15, 256–266.
- Kellogg, R. T. (1994). *The psychology of writing*. New York: Oxford University Press.

- Kluwe, R. H. (1982). Cognitive knowledge and executive control: Metacognition. In D. R. Griffin (Ed.), *Animal mind—human mind* (pp. 201–224). New York: Springer-Verlag.
- Kuhn, T. S. (1970). *The structure of scientific revolutions*, 2nd ed., enlarged. Chicago: The University of Chicago Press.
- Markus, H. & Nurius, P. (1986). Possible selves. *American Psychologist*, 41, 954–969.
- McConkie, G. W., Hogaboam, T. W., Wolverton, G. S., Zola, D. W., & Lucas, P. A. (1979). Toward the use of eye movements in the study of language processing. *Discourse Processes*, 2, 157–177.
- McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 115–130). New York: Guilford.
- Montgomery, H. & Svenson, O. (1989). *Process and structure in human decision making*. Chichester, Great Britain: John Wiley & Sons.
- Moore, E. C. (Ed.). (1972). *Charles S. Peirce: The essential writings*. New York: Harper & Row.
- Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, 26, 125–141.
- Newell, A. & Simon, H. A. (1973). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Nystrand, M. (1982). Rhetoric's audience and linguistics speech community: Implications for understanding writing, reading, and text. In M. Nystrand (Ed.), *What writers know: The language process, and structure of written discourse* (pp. 1–28). New York: Academic Press.
- Nystrand, M. (1989). A social-interactive model of writing. *Written Communication*, 6, 66–85.
- Nystrand, M. (2006). The social and historical context for writing research. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 11–27). New York: Guilford.
- Nystrand, M. & Himley, M. (2001). Written text as social interaction. *Theory into Practice*, 23, 198–207.
- Plato (1961). *The collected dialogues of Plato*. New York: Bollingen Foundation.
- Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77, 97–131.
- Pressley, M. & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Prior, P. (1991). Contextualizing writing and response in a graduate seminar. *Written Communication*, 8, 267–310.
- Prior, P. (2006). A sociocultural theory of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 54–66). New York: Guilford.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Reder, L. M. & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Rijlaarsdam, G. & van den Bergh, H. (1996). The dynamics of composing—An agenda for research into an interactive compensatory model of writing: Many questions, some answers. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theory, methods, individual differences, and applications* (pp. 107–126). Mahwah, NJ: Lawrence Erlbaum Associate.
- Sapir, E. (1961). *Culture, language, and personality*. Berkeley, CA: University of California Press.
- Shaughnessy, M. (1977). *Errors and expectations*. London: Oxford University Press.
- Vygotsky, L. (1991). *Thought and language*. Cambridge, MA: The MIT Press.
- Witte, S. P. (1992). Context, text, intertext: Toward a constructivist semiotic of writing. *Written Communication*, 9, 237–30.

Part V

Science and Mathematics

10 The Interplay of Scientific Inquiry and Metacognition

More than a Marriage of Convenience

Barbara White

University of California at Berkeley

John Frederiksen

University of Washington

Allan Collins

Northwestern University

The scientific enterprise is a form of collaborative learning that enables society to develop knowledge about the world—knowledge that is useful for predicting, controlling, and explaining what happens as events occur. Creating scientific communities in classrooms, by engaging young learners in theory-based empirical research, is a highly challenging yet important educational goal (Anderson, 2002; Blumenfeld, Soloway, Marx, Krajcik, Guzdial, & Palincsar, 1991; National Research Council, 1996, 2007; Schraw, Crippen, & Hartley, 2006). To achieve this goal, and enable students to learn about the nature and practices of scientific inquiry, the development of metacognitive knowledge and capabilities is crucial.

In this chapter we outline the metacognitive expertise that is needed to understand and regulate inquiry processes as students undertake research projects, and we elaborate why developing this type of expertise is important. We present evidence that students' learning of scientific inquiry can be enhanced by providing them with explicit models of inquiry goals and strategies, while also teaching them self-regulatory processes. This approach is particularly effective for lower-achieving students.

Explicit models of inquiry and metacognitive expertise can be embodied in educational software as advisors, like the Theorizer and Reflector, which define and promote effective inquiry processes. For example, advisors can suggest appropriate goals and strategies to pursue, give examples of strategies in action, and provide criteria for monitoring their effectiveness. In addition to guiding students as they undertake research projects, the software advisors can be adopted, tested, and modified. For example, students can take on the role of an advisor in their research groups, so that one student becomes the Theory Manager, another the Reflection Manager, and so forth. Furthermore, students can undertake research projects in which they create and test competing models of their expertise. They may, for instance, compare different strategies for generating theories or for reflecting on their inquiry processes. In this way, students invent and test their own conceptions of how to do various aspects of inquiry, how to learn through inquiry, and how to monitor and reflect on their progress. We provide evidence of the pedagogical utility of such role playing and “inquiry about inquiry,” which engages students in the recursive application of scientific inquiry and metacognition to foster the development of both types of expertise.

Metacognition and Scientific Inquiry

Ann Brown (1987) identifies two basic types of metacognitive expertise. The first is knowledge about cognition, which could be called self-understanding. It includes knowing what you know and don't know, as well as how you learn, through processes like scientific inquiry, to increase your understanding of the world. The second type of metacognitive expertise is related to managing and improving your cognition and is often called self-regulation. It includes planning, monitoring, and reflecting, which includes being able to plan a research project, monitor your progress, and think about how you could do better next time.

Knowing What You Know and Don't Know

Engaging in successful scientific inquiry requires developing metacognitive capabilities, such as being able to use conceptual models to generate explanations and check for understanding (Chi, Bassok, Lewis, Reimann, & Glaser, 1989). Engaging in such practices not only facilitates students' learning of inquiry, it also fosters their metacognitive development. If students are going to become proficient at scientific inquiry and learn how to understand a domain deeply, it is important that they be able to recognize gaps and inconsistencies in their conceptual models for understanding a particular domain. We have been struck by the finding that some students simply grab onto the first idea that occurs to them and stick with it, while others recognize the limitations in their understanding and pursue alternative ideas. Furthermore, figuring out how different ideas fit together, and whether in fact they do fit together, is critical. Unless students learn to recognize anomalies and contradictions, and are motivated to pursue them deeply enough to resolve such discrepancies, it is unlikely that they will be able to engage in successful scientific inquiry and build coherent conceptual models of complex domains.

Knowing How You Learn Through Inquiry

It is also important that students develop meta-knowledge about inquiry itself. Knowing about the nature and practices of scientific inquiry fosters proficiency in doing science, as well as in appreciating the role of science in the evolution of our societies (National Research Council, 2007). Such meta-level expertise includes knowing about the nature of scientific theories and questions, understanding the different types of investigations and data analyses that are possible, and appreciating the role that each type of investigation and analysis can play in testing hypotheses and in developing and refining models and theories.*

* Like many words in the English language, the terms "model" and "theory" have multiple meanings. When we use the term "model" or "conceptual model" as a noun, we mean a representation of a system that characterizes its structure or behavior, which can include rules, representations, and reasoning structures that allow one to predict and explain what happens as events occur. The scientific use of the word "theory" is similar to our definition of the term "model," although theories may be thought of as more comprehensive and models as more specific. In this chapter, we argue that the goal for scientists and students is to create comprehensive theories, which include components of different types (such as concepts, predictive laws, and explanatory models), as well as interlinked models that embody different perspectives (such as macroscopic and microscopic). However, we also use the term "theory" to refer to initial, tentative ideas about the nature and causes of phenomena, particularly when we are referring to the theories that students are creating and testing. These typically include key variables (concepts), predictions, and explanations, but they are not necessarily comprehensive theories that have evolved to encompass and integrate different perspectives.

Managing Your Cognition

To engage in successful inquiry, it is also essential that students learn planning, monitoring, and reflection techniques (Azevedo & Cromley, 2004; Schunk & Zimmerman, 1998; Zimmerman, & Schunk, 2001). Careful planning, for example, is crucial to many aspects of scientific inquiry, ranging from designing an investigation to determining how to analyze the resulting data. In designing an experiment, for instance, students must determine all of the steps that need to be completed. As part of their planning process, they need to learn to envision the possible outcomes for their experiment to make sure that it will enable them to determine which hypotheses are supported and which are not.

As learners carry out scientific inquiries, they need to monitor a range of things, including inquiry processes and products, progress at achieving goals, and metacognition itself. These can be monitored by getting students to ask themselves a variety of questions. For instance, Schoenfeld (1987) has students ask themselves three high-level metacognitive questions: (a) what are we trying to do; (b) why are we trying to do that; and (c) are we making progress? One can also ask students to monitor their achievement of widely applicable cognitive, social, and metacognitive goals, such as: Are we (a) reasoning carefully, (b) collaborating effectively, and (c) monitoring our progress sufficiently? It is also important for students to learn to monitor and evaluate their work using more focused criteria that characterize good work for each aspect of inquiry. For example, such self-assessment, in the context of arguing for a particular scientific explanation after doing an investigation, might include asking: (a) does our explanation address our research question; (b) does it seem convincing; and (c) does it fit the experimental findings? There are a large number of such monitoring questions that students should ask themselves as they engage in scientific inquiry, and doing so plays a role in carrying out good research and in coming to understand its nature and purpose.

Improving Your Cognition

Reflection can be used as an opportunity to improve the processes of inquiry, including the metacognitive practices it requires. Engaging in collaborative reflection, after students have completed a research project for example, can get them to think about how they could improve their processes for generating competing hypotheses or their processes for planning an investigation to test their hypotheses. In this way, students can become reflective practitioners (Schön, 1983) of scientific inquiry and gradually, over time, improve their inquiry practices and their understanding of their goals and purposes.

Metacognition and the Teaching of Scientific Inquiry

Unfortunately, not enough emphasis is placed on the development of metacognitive practices in the national curricular standards for science education. The standards put forward by the National Research Council (1996), for example, emphasize developing knowledge of a domain through inquiry, but they place far less emphasis on the need for metacognitive knowledge and skills. Yet increasingly, research has shown that metacognitive expertise is needed in developing knowledge through inquiry (Chinn & Malhotra, 2002; Frederiksen & White, 1997; Georghiades, 2004; Hogan, 1999; Kuhn & Pearsall, 1998; White & Frederiksen, 1998), and is critical in transferring one's capabilities for learning in one domain to learning in new domains, as well as taking charge of one's own learning (Bransford & Schwartz, 1999; Scardamalia & Bereiter, 1991). There is also evidence that feelings of self-efficacy in learning play a strong role in students' motivation

and interest in learning (Bandura, 1997; Pintrich & de Groot, 1990; Schunk & Schwartz, 1993). Building metacognitive knowledge of oneself as a learner contributes to viewing oneself as an able learner (i.e., it develops self-efficacy), which influences not only success in learning, but also motivation to learn (Brown, 1988; Corno, 1986; Zimmerman, 1989; Zimmerman & Schunk, 2001).

Our basic thesis is that to become an effective inquirer, a person must develop the various types of metacognitive knowledge and capabilities that we outlined above. Science education needs to take metacognition seriously in order to educate students effectively (Baird, 1986; Baird & White, 1996; Schraw, Crippen, & Hartley, 2006; White & Gunstone, 1989; Yore, Pimm, & Tuan, 2007). To date, an emphasis on students' metacognitive development has been largely missing from the teaching of science in the vast majority of classrooms.

In recent years, however, educational researchers have been pursuing various approaches to addressing this need. For example, there have been a number of attempts to promote metacognition through computer scaffolding as students try to generate and refine scientific explanations (e.g., Graesser, McNamara, & VanLehn, 2005; Quintana, Zhang, & Krajcik, 2005; Sandoval & Reiser, 2004; Schwartz et al., this volume). Other researchers have investigated the utility of providing prompts, and other devices, to foster students' reflection and self-regulation as they work on science projects (e.g., Davis, 2003; Loh, Radinsky, Reiser, Gomez, Edelson, & Russell, 1997; Toth, Suthers, & Lesgold, 2002). Some researchers have tried to make the nature of scientific theories more explicit to students (e.g., Lehrer & Schauble, 2005; Perkins & Grotzer, 2005; Schwarz & White, 2005; Slotta & Chi, 2006), while others have emphasized making inquiry processes more explicit (e.g., Blank, 2000; Linn, Davis, & Bell, 2004; Metz, 2004; Schauble, Glaser, Duschl, Schultz, & John, 1995; Smith, Maclin, Houghton, & Hennessey, 2000). Finally, some researchers have fostered students' understanding of the scientific enterprise by employing a "community of learners" approach, which emphasizes collaborative inquiry and knowledge building (e.g., Brown & Campione, 1996; Herrenkohl, Palincsar, Dewater, & Kawasaki, 1999; Hogan, 1999; Metz, 2000; Scardamalia & Bereiter, 1994). Such approaches can be highly successful at developing students' understanding of science, thereby meeting important goals for science education, while also symbiotically fostering students' metacognitive development.

In what follows, we outline some of the different types of meta-level knowledge and capabilities needed to understand and do scientific inquiry. Then we present examples of how such meta-level expertise can be fostered and utilized in science teaching, and explain, in more depth, why we think metacognition should be an integral component of science education.

Meta-Knowledge of Scientific Inquiry

Scientific inquiry can be viewed as a process of oscillating between theory and evidence, in a practice of competitive argumentation that leads to the development, testing, and elaboration of scientific laws, models, and theories. The ultimate goal is to create alternative theories and develop arguments that employ explanations and evidence to support or refute those theories (cf., Carey & Smith, 1993; Driver, Newton, & Osborne, 2000; Duschl, 2007; Duschl & Osborne, 2002; Kuhn, 1993; National Research Council, 2007). The transition from making theories to seeking evidence, through an investigation, is one where the generation of questions and hypotheses derived from theory is crucial. The transition from carrying out an investigation to the refinement of a theory is one in which data analysis and synthesis are central. This view leads to a basic model of

scientific inquiry that has four primary processes: (1) Theorizing, (2) Questioning and Hypothesizing, (3) Investigating, and (4) Analyzing and Synthesizing. Associated with each of these primary processes is a regulatory process that monitors how well the process is being carried out and whether another process should be invoked to deal with issues that arise.

In our earlier work on teaching scientific inquiry to young learners, we portrayed such a model as an Inquiry Cycle, which provides a scaffold for inquiry in the form of a series of steps that one undertakes in a never-ending cyclical process of generating, testing, and elaborating scientific principles and models, with the ultimate goal of developing a widely useful, accurate, and comprehensive theory for a given domain. This is, of course, a simplified view: mature scientific inquiry does not necessarily proceed in this step-wise fashion. For one thing, it is possible to start anywhere in the sequence. So, for example, one might start with vague questions that are not based on a particular theory, or one might start with an investigation or with existing data to generate theoretical ideas. Furthermore, one does not necessarily proceed through these “steps” in order. For instance, analyzing data can lead to the need to do further investigation. So the critical elements in the scientific enterprise are closely intertwined and any view of science education that underplays one of these elements fundamentally misleads students as to the nature of science (Chinn & Malhotra, 2002). Nonetheless, for pedagogical purposes, presenting students with an inquiry cycle, in which one starts with theorizing and questioning, is an effective initial model that can enable students to develop capabilities for inquiry, as well as an understanding of its constituent processes (White & Frederiksen, 1998, 2005a, 2005b).

In the next four sections, we briefly describe meta-scientific knowledge in terms of its four components: meta-theoretic knowledge, meta-questioning knowledge, meta-investigation knowledge, and meta-analysis knowledge (note that we do not mean “meta-analysis” in the statistical sense). We should emphasize though that meta-questioning knowledge includes meta-knowledge about forming both research questions and hypotheses, while meta-analysis knowledge includes meta-knowledge about data analysis, synthesis, and argumentation.

This model of scientific inquiry reflects the way most sciences include two camps: the theoreticians and the empiricists. Theory and empirical investigation form the two poles of science. Research questions form a bridge between these two poles, in which competing theories generate alternative hypotheses that are tested through empirical investigation. Analysis and synthesis form the other bridge between the poles by providing ways to represent and interpret data from the investigation to bear on the theories in competition.

Meta-Theoretic Knowledge

Meta-theoretic knowledge includes knowledge about the nature of scientific models and theories. In their work on epistemic forms and games, Collins & Ferguson (1993) characterized three types of models (or epistemic forms) that researchers use to guide their inquiry: structural, causal (or functional), and dynamic (or process) models. The different forms of structural models include hierarchies, cross-product tables (e.g., the periodic table), stage models, primitive elements (e.g., chemical elements), and comparison tables. Similarly there are different types of causal models, such as causal chains, form-function analysis, and multifactor models (as in medicine). Finally there are different dynamic model types, such as system-dynamics models, production systems (situation-action rules), and agent models. All of these representational forms have epistemic games (i.e., rules and

strategies) associated with them, which are practices scientists use as they construct models to characterize and theorize about different phenomena.

Different model types serve different purposes. Structural models highlight the relationships between different elements in the models. Causal models depict the causal and functional dependencies between elements in the models. Dynamic models allow one to “run” models to see the consequences of different assumptions and principles embodied in the models. These runnable models can unpack mechanisms that explain the causal relationships depicted in static causal models.

Scientific theories in our view are made up of a number of linked models. In chemistry, for example, the primitive elements (hydrogen, helium, etc.) are arranged in a cross-product table (i.e., the periodic table). There is an underlying atomic structural model of protons, neutrons, and electrons arranged in shells that accounts for the structure of the periodic table. There are also constraints that determine how different elements combine into molecules, based on their atomic structure. Hence the standard theory in chemistry is made up of different types of models linked together in systematic ways.

Frederiksen and White (2002; Frederiksen, White, & Gutwill, 1999) have shown how different models of electricity are linked together, in particular how circuit diagrams, constraint equations, local-flow models, and four other model types capture different aspects of the behavior of electrical systems. These models are linked in three different ways. First of all, models are linked vertically when the behavior of one model type is derivative or emergent from the behavior of another model type (Wilensky & Resnick, 1999). For example, Frederiksen and White (2002) describe how the behavior of constraint systems can be derived from the local-flow model of electricity they present. Secondly, models are linked horizontally when there is a “progression of models” (White & Frederiksen, 1990), such that a higher-order model is derived from a lower-order model by adding new rules or entities. For example, a simple local-flow model of circuit behavior allows people to solve problems about serial circuits, and an elaboration of the model allows people to solve problems about serial, parallel, and hybrid circuits. Finally, models are linked in a coordinated way when the models represent distinct characteristics, or perspectives, of a system in compatible ways (White & Frederiksen, 1990). For example, circuit diagrams and constraint equations represent different aspects of an electric circuit, but the two representations for any given circuit must be coordinated.

Given the growing importance of modeling in science education (Gilbert, 1991; Halloun, 2004; Hestenes, 1987; Lehrer & Schauble, 2000; Mellar, Bliss, Boohan, Ogborn, & Tompsett, 1994; Schwarz & White, 2005; Smith, Snir, & Grosslight, 1992; Stewart, Cartier, & Passmore, 2005; White, 1993; Windschitl, Thompson, & Braaten, 2008), we think it is critical that students learn about the different forms that models can take and how different models can be linked together to form a coherent and powerful theory. Hence, the essential meta-theoretic knowledge that people need to learn is how theories and models are created, refined, and coordinated. We think these various pieces fit together to form the basis for the development of meta-theoretic knowledge in science.

Meta-Questioning Knowledge

In order to evaluate theories, it is necessary to turn elements of the theories into research questions that can be directly investigated. Sometimes research questions are quite vague (e.g., What are the precursors to heart disease?) and sometimes the questions are specific (e.g., Does taking a particular drug reduce one’s cholesterol level?). The hypotheses in any study are the different possible answers to a research question based on alternative theoretical positions or assumptions. Ideally, the different answers and their implications

should be specified in advance of the investigation. However, if a research question is vague, this may not be possible.

The different epistemic forms (i.e., types of models) generate different types of research questions. For example, structural models generate questions such as: What are all the different types of X, and what stages does X go through as it evolves? Causal models generate questions such as: Does Y cause X? or What are all the factors that affect X? Dynamic models generate questions such as: What process produces X? and What are the rules of interaction between X and Y? These are some of the most common research questions that arise as scientists create models to express and develop their theoretical ideas.

Ideally research questions help to differentiate between possible theories. Finding crucial questions that in fact distinguish between alternative theories is very difficult. And when a crucial question is investigated, the researchers whose theories are not supported by the data usually can come up with some explanation that still preserves their theory, albeit in a modified form. But even in investigations that do not compare alternative theories, researchers have to come up with questions and data that enable them to differentiate their explanations for their findings from obvious alternative explanations that other researchers might generate.

When one research question is answered, it often raises a set of related questions. One way this occurs is when a particular structural pattern is found, as when Mendeleev discovered the periodic table, it raised the question of why the elements in a single column have similar properties. This question led eventually to the model of the atomic structure of atoms. Similarly a causal model, such as a multifactor model, raises questions about the mechanisms that lead each of the factors to have the given effects. These two examples show how answering one research question can lead researchers to generate related questions about underlying processes, structures, and mechanisms.

The meta-questioning knowledge that students need to acquire includes learning about the different types of research questions that can be asked, and how each type of question is related to particular epistemic forms. Students also need to develop an understanding of how questions can be created to distinguish between competing theories and how, in the process of creating a deeper, more coherent theory for a domain, one question leads to another.

Meta-Investigation Knowledge

The third element of scientific meta-knowledge is an understanding of the different forms that scientific investigations can take. There are many different investigation methods, but they generally fall into two basic types: (1) exploratory inductive investigations (often referred to as scientific induction) and (2) confirmatory investigations (often referred to as the hypothetico-deductive method). Exploratory inductive investigations are employed when one has broad research questions and some general theoretical ideas that suggest interesting data sources to study, but which are not specific enough to generate particular hypotheses. The goal is to obtain data that will constrain one's efforts to develop more detailed theories. Confirmatory methods are used when one has a well-developed theory or set of theories in mind, which allow one to develop a set of theoretical hypotheses to test. The goal is to test each of the hypotheses to see if the findings are consistent with the theoretical predictions. This allows one to determine which theory or theories are most consistent with the data and which are not suitable for explaining the phenomena that have been investigated.

Inductive Investigations

Galileo is famous for developing exploratory inductive methods in science. In his experiments on pendulums and gravity, he systematically varied the elements that he thought might affect the period of the pendulum and the speed of a ball rolling down an incline. From these exploratory investigations, he derived equations for the motion of pendulums and falling bodies. The Framingham Heart Study is a modern variation on his method, using natural variation rather than controlled manipulation. The investigators in this study collected data from many people in Framingham, Massachusetts, on a large number of variables that they thought might influence the likelihood of getting heart disease. They then followed the people over many years to see if they developed heart disease and identified a number of variables that were precursors to heart disease.

In exploratory inductive investigations, different methods and data sources are often used to cover the phenomena of interest in order to construct a more robust theory. Dewey (1910, Chapter 7) suggested three principles for regulating the observation and collection of data in forming “explanatory conceptions or theories” that seem to be wise advice today: (1) Care must be taken in differentiating between what is observed and what is inferred, so as not to jump to hasty conclusions about one’s theory. (2) One needs to look for multiple cases to see how general one’s conclusion is, but one also needs to look for contrasting cases in order to determine the factors that are critical to the conclusion. (3) One needs to look for exceptions and contrasts that may challenge one’s initial conclusions and suggest others. One often learns more from examining anomalous cases that don’t have the expected features. (This principle is similar to having control conditions in a confirmatory investigation or experiment.)

Even when following these principles, a limitation of such exploratory investigations is that the researchers may still have systematic biases in what data are collected and in how they are interpreted. This is why confirmatory studies are critical to the development and refinement of scientific models and theories. In carrying out exploratory research, it is important to be aware of how one employs one’s theories in selecting and interpreting data. Confirmatory investigations that test hypotheses are needed to evaluate the accuracy of a theory’s predictions and to resolve conflicts between different theories.

Confirmatory Investigations

Confirmatory investigations, which are designed to test theoretical hypotheses, can take many different forms. The most common is the randomized controlled trial, in which one or more proposed hypotheses are tested by comparing conditions that correspond to each of the hypotheses one is testing. Often such a test of a hypothesis contrasts an “experimental” condition, which includes some particular feature, with a control condition that lacks that feature. In order to ensure the generality of the findings, multiple experimental units (test trials, subjects tested) are assigned randomly to different conditions and then compared to see if differences in a dependent variable are consistent with what was predicted by the hypothesis. Often special efforts are made for controlling variables that one thinks might also affect the results. Confirmatory studies are also critical in determining the boundary conditions for a theory—that is, the range of situations over which a theory applies.

Studies designed to confirm or test hypotheses are best suited for situations where there are a small number of hypotheses and variables. When situations are complex, investigators may only be able to test a few specific predictions of a theory. Many confirmatory studies simplify or standardize the situation and collect a large amount of data, hoping

that factors that have not been controlled are contributing randomly to the effects being investigated and will not affect the group averages in any systematic way. Another way to deal with complexity is to use multivariate methods, such as regression or covariance analyses. The intent in these methods is to control complexity by taking into account, through statistical adjustments, factors other than those specified in the hypotheses that might have effects on the results.

We have argued that exploratory inductive studies help in constructing theories, whereas confirmatory investigations are used to test hypotheses that represent competing theories; however, the process is really cyclical. Often confirmatory studies, especially if they include collecting rich data, provide clues for further theory development through an embedded inductive investigation of those data. This can lead to theory refinement and suggest new confirmatory investigations that can further refine and test the theory. Thus in scientific inquiry, testing hypotheses deduced from theories, and interpreting patterns in data to construct new theories, are intertwined. This process is complex and depends on meta-knowledge of the forms of inquiry one is engaging in at a given time, and how they are interrelated as one moves from one form of investigation to another. Meta-investigational knowledge also makes one aware of the pitfalls and limitations of the forms of investigation one is using at any particular time.

Meta-Knowledge for Data Analysis

The main purpose of data analysis is to support the development of convincing arguments, which show how the findings from an investigation support particular conclusions and have implications for theories. Data analyses are systematic procedures for examining the information obtained in an investigation. These procedures can best be understood by the research functions they support: (a) the representation of data, (b) the confirmation or refutation of existing hypotheses, (c) the induction of new hypotheses, and (d) generalization.

Representing Data to Reveal Patterns of Theoretical Interest

Following the collection of data in an investigation, one needs to create useful data representations, which code and display data in ways that (a) reveal if patterns that are predicted by a theory are present, or (b) reveal new patterns that may require modifications or additions to the theory. In order to see patterns, data obtained in different situations need to be represented using similar measurements or coded qualitatively using similar categories. This makes it possible to make comparisons of data across different situations.

There are many kinds of patterns in data, paralleling the many kinds of relations among factors or variables that are generated by theories or models. In complex, multi-variable data sets, the number of pair-wise relations among variables can be great, and the possibility of interactions among variables increases the number of possible patterns even further. Experience in reading other investigators' data analyses will lead to creating a "library" of forms for displaying data (cf., Giere, 1991). For instance, a graph of average values obtained before and after a treatment, shown for two different treatments, may be a good way to see if there is an interaction between treatment and effect.

Exhaustively searching for meaningful relationships is often impractical. The particular tools one uses in data analysis to reveal patterns are often guided by the epistemic forms of the models one has in mind. Here choices of epistemic forms and data analysis methods are entangled: you see data patterns through analyses that are themselves suggested by theories. In arguing for a particular interpretation of data, you need to be

cognizant of how other investigators with different theoretical orientations might interpret the data. One of the most difficult things in data analysis is to be able to “put on the hat” of a different theorist and consider alternative forms of analysis which that might entail.

Confirming (or Disconfirming) Hypotheses

To test hypotheses, one needs to obtain evidence of the soundness of each competing hypothesis. Patterns found in data provide evidence for whether a hypothesis is confirmed or disconfirmed. Confirming a hypothesis increases confidence in the theory’s accuracy, but does not confirm the theory itself, because other theories might be constructed that lead to the same prediction and hypothesis. However, disconfirming a hypothesis can support an argument for rejecting the associated theory. Popper (1963) argues that strong theories are subject to refutation when tested, but can never be fully confirmed. Theories that are not fully specified are hard to disconfirm, because they can be augmented to account for factors that had been left out. Having meta-knowledge of the theoretical status of confirming and disconfirming evidence for a theory should help investigators to be careful in making inferences based on findings. It also motivates investigators to consider alternative hypotheses when they are designing their investigation.

Induction: Exploring the Data to Uncover New or Unanticipated Phenomena

Often results reveal patterns that suggest new relationships among variables or factors that have interesting theoretical interpretations. If such a pattern is first noticed in a subset of the data (for example, in studying a particular case), one seeks to verify it by exploring its presence in other sets of data that are available. Finding unanticipated phenomena or relationships among variables may provide further support for one’s current theory, or suggest ways to augment the theory, or require the invention of new theories. In searching for interpretable patterns, meta-theoretic knowledge can be very useful. For instance, one might think of a particular epistemic form, such as stage theory, as a way to look for patterns (Collins & Ferguson, 1993).

Generalizability: Establishing the Generality of the Findings

Theories are expressions of relationships that have general applicability across a range of situations. In scientific inquiry, establishing the generality of a theory is important. Commonly, one obtains data from a sample of different situations, or individuals, to provide evidence for the consistency of the results that are predicted by the theory across the range of circumstances to which it purportedly applies. However, a theoretical argument for the generalizability of a conclusion to other situations can also be made. This can take the form of specifying the conditions that are necessary for a conclusion to apply, with the implicit suggestion that other factors not mentioned are irrelevant.

All of these functions of data analysis are applicable and important, whether or not the data that are collected are quantitative measures of variables, detailed qualitative recordings or field notes, or a mixture. Data collected should allow researchers to test specific hypotheses, while at the same time supporting in-depth studies of the underlying processes. This is why many scientists keep extensive laboratory or field notes when they are conducting their investigation. Displaying data using multiple forms makes it possible to see different patterns. Ideally the researchers should be able to synthesize findings of all

their analyses to produce a coherent interpretation, or argument, that supports a particular theory, one which provides a better account than other, competing theories.

Metacognitive Control of Inquiry Processes

The inquiry process is complex. In carrying out research, one draws upon meta-knowledge of the top-level structure of inquiry and of its constituent processes: knowing about the forms that theories can take, developing questions and hypotheses that can test implications of theories, managing multiple goals and strategies for designing an investigation, analyzing data, and so forth. It is clear that inquiry involves the interplay of a large repertoire of processes, each with its underlying goals and strategies. To manage this system, one needs to understand when, why, and how the various processes are engaged in the course of carrying out an investigation, what are the products that are created by each process, and how those products are used as “inputs” for other processes. One also needs to have metacognitive expertise for controlling and improving inquiry, that is, knowledge and capabilities for self-regulation.

Zimmerman (1998) describes self-regulation as having three major phases: (a) forethought, (b) performance or vocational control, and (c) self-reflection. Forethought includes goal setting and strategic planning. Performance control has two aspects: (1) self-control, which includes deciding how to proceed in attending to and carrying out a task, and (2) self-monitoring, which is judging how well your processes for achieving the task’s goals are working. This helps you decide whether to try a different way to achieve the goal, and whether the products of your work are good enough to move on to another inquiry task or goal. Self-reflection refers to evaluating one’s performance of inquiry tasks, using particular criteria for each task, in order to identify ways to improve one’s inquiry products and processes. By choosing criteria carefully, students’ reflection can be directed towards developing inquiry meta-knowledge, which includes understanding the goals and purposes of the various inquiry processes as well as how they are interrelated. Self-reflection is useful as a basis for self-monitoring at all levels in the inquiry process, for instance in judging the quality of the design for your investigation, or in considering how well you have accomplished a sub-goal, such as choosing dependent and independent variables. This means that the self-regulatory system is used recursively throughout the sequential and hierarchical levels of inquiry (cf., Winne & Hadwin, 1998).

In general, we believe that for students to engage in self-regulated inquiry, they not only need to be provided with a means for developing the meta-knowledge of inquiry we have described, they also need to have explicit performance standards for each of the goal-directed processes in which they are engaged. Reflection is therefore central to both understanding and managing the movement through this network of inquiry processes. Participants in a classroom research community need to be involved explicitly in a reflective process in which they review their processes of working and the products of their investigation at every stage of their work. This reflective process should be a social one, so that students may see how multiple perspectives can be applied in viewing one’s own and others’ work as they carry out the processes of inquiry. This social process allows students to practice and internalize habits of reflection (Vygotsky, 1978).

The importance of such metacognitive behavior has been emphasized by researchers (e.g., Brown & Campione, 1996), who maintain that monitoring and reflecting on the process and products of one’s own learning are crucial to successful learning as well as to “learning how to learn.” Research on learning shows that many students, particularly lower-achieving students, have inadequate metacognitive processes and their learning suffers accordingly (Campione, 1984; Chi et al., 1989; Zohar & Dori, 2003). Thus,

if you introduce and support such reflective, metacognitive processes in the curriculum, the students' learning and inquiry should improve, at the same time as they develop self-regulatory capabilities and meta-knowledge about scientific inquiry.

Approaches to Developing Students' Meta-Knowledge in Science Education

There is strong evidence that students and many adults have great difficulty in understanding the nature and relations among theories, hypotheses, and evidence (Carey & Smith, 1993; Smith & Wenk, 2006). This is viewed as a major handicap in understanding the nature of science, and it is also critical knowledge for carrying out scientific inquiry. Yet there is also evidence that young people *can* develop an understanding of the role that theories play in investigating and developing an understanding of the world (Smith et al., 2000). We believe that by bringing meta-knowledge about science and reflective metacognitive practices together within a rich learning environment, students will be able to develop these kinds of knowledge and capabilities and use them in inquiry.

There are a variety of ways to foster metacognitive knowledge and capabilities in students. Metacognitive thinking can be modeled so that students can see metacognition in action (Schoenfeld, 1983; 1987). Teachers can encourage and scaffold metacognitive practices. A common technique is to prompt students to do a metacognitive task, such as evaluating whether they have identified all possible outcomes to an experiment. Getting students to take time to reflect is also a productive way to encourage students to be more metacognitive. Sometimes metacognitive thinking is promoted and scaffolded by a metacognitive tool, such as a research journal, to prompt students to plan, monitor, and reflect on their work. In addition, as we will argue later, getting students to invent and investigate metacognitive processes themselves may be an effective way to foster metacognitive thinking and development. In all of these approaches, the aim is to introduce metacognitive practices into the culture of the classroom. This can include engaging students in collaborative planning, peer assessment, and collective reflection as they work together on research projects. In this way, students and teachers can model and support a range of metacognitive processes that facilitate scientific inquiry and learning. We argue that each of these methods can do this effectively if, in some way, it provides meta-theoretic information and insights about the inquiry processes in which students are engaged.

Learning by Studying a Scientist's Notebook

One example of a metacognitive approach is Magnusson and Palincsar's (2004) design and study of an innovative form of science text, which should help students in learning inquiry science while also promoting literacy. They first carried out an analysis, which revealed many parallels between the processes used in constructing knowledge in text comprehension and those used in learning through inquiry. However, they also found that that texts used in science instruction are more concerned with the presentation of information rather than with how scientific information is developed. In response to this mismatch, they developed a new type of text that is intended for use in inquiry-based science classes. It illustrates both the processes of scientific reasoning used in scientific inquiry, and the metacognitive skills needed to enlist appropriate learning strategies and monitor their success in constructing meaning.

The text is modeled after the notebook of a fictitious scientist, named Lesley, who uses her entries to (a) identify the phenomenon she is investigating, (b) think aloud about how she can accurately model the phenomenon for the purposes of investigation, (c) make decisions

about how she will most effectively represent the data that she is collecting, (d) share her data and the claims that she believes she can make from these data, (e) respond to the critical reactions of her colleagues as they weigh the evidence for her claims, and (f) revise her thinking as she gathers new data or considers alternative explanations. These kinds of entries in the notebook represent important forms of scientific meta-knowledge. For example, the notebook illustrates metacognitive practices that are used in carrying out inquiry through Lesley's evaluations and revisions during the course of her investigations. In one illustration, Lesley has identified a set of claims based upon her preliminary evidence. Subsequently, with the urging of her peers, she collects more precise data, which lead her to revise her initial claim. This kind of modeling of a scientist's metacognitive thinking shows learners how to monitor and reflect on their work as they carry out scientific investigations. The notebook also illustrates the social nature of scientific deliberation.

Magnusson and Palincsar carried out studies showing how using notebook texts, along with carrying out inquiry, facilitates students developing scientific concepts of knowledge construction, while also gaining a "critical stance" towards text comprehension and metacognitive capabilities, such as checking their sense-making and choosing strategies to improve their understanding. For example, in one class, students investigated how light interacts with objects while they were also reading Lesley's notebook, and they compared their findings with Lesley's findings. This led them to identify variables that might be different in Lesley's and their investigations and that might have caused the different outcomes. The notebook texts also lead the students to reason about the design of their investigations, their measurement procedures, and the relations of claims and evidence.

Learning Using Guided Inquiry and Reflective Assessment

Another approach is to provide students with research notebooks that structure their work in carrying out an inquiry project, and to have students reflect on their work as they proceed. This is the approach we took in our earlier research with the ThinkerTools Inquiry Curriculum (White & Frederiksen, 1998) in which we first studied the effect of metacognitive reflection on students' learning while they are engaging in collaborative inquiry. The research notebooks present a model that portrays the structure of the inquiry process, and incorporates self-assessments for each inquiry step to foster reflection. Students formulated and tested models of force-and-motion phenomena as they carried out a series of seven investigations. In each investigation (i.e., instructional module), students experienced all of the stages of inquiry (Question, Predict, Experiment, Model, Apply), while their overall research goal was to work to extend their theory of force and motion to cover more complex phenomena.

In this study, we had matched experimental and control classes for each of three teachers, so that we could compare their students' learning. In the experimental or "Reflective Assessment" classes, metacognitive reflection was introduced by having students assess their work using a set of criteria that represent high-level cognitive and social goals, such as reasoning carefully and collaborating effectively. When students evaluated the research they had just completed, they were asked to write a justification for their score. Students in the control classes did not engage in reflective assessment, but spent an equivalent time commenting, at the end of each module, on what they did and did not like about that module, as opposed to the experimental classes who reflected, as described above, on their own research processes and products.

Students carried out two independent research projects, one about halfway through the curriculum and the other at the end (these were less scaffolded than the other five modules of the curriculum). To study the effects of students' prior academic background, we

compared students who were in the lowest, middle, and highest thirds in their scores on a standardized test (the CTBS) used in the school districts. For the sake of brevity, we have added the scores for the two research projects together in showing our findings (see Figure 10.1). These results reveal that students in the reflective assessment classes have higher-rated research projects than students in the control classes. The results also show that reflective assessment is particularly beneficial for the lower-achieving students. The differences among mean project scores for the reflective assessment group are statistically indistinguishable for the three levels of CTBS scores, while those for the control group differ significantly.

We also tested students' understanding of inquiry using an individually administered written assessment called the Inquiry Test. This test asks students to engage in a thought experiment. It provides them with a research question (i.e., what is the effect of varying the weight of an object on what sliding friction does to its motion) and asks them to do the following: (a) generate and justify two alternative hypotheses about the answer to the question, (b) design an experiment to test these hypotheses, (c) make up data that are consistent with their design, (d) analyze their made-up data and reach a conclusion, and (e) explain which hypothesis, if any, is supported by their research. The Inquiry Test was scored solely with regard to inquiry skills. In other words, students were assessed on their understanding of the inquiry processes not for their conceptual understanding of physics.

This Inquiry Test was given to students before and after the ThinkerTools Inquiry Curriculum was completed. Comparing these scores allowed us to see if the gains in inquiry knowledge were greater for students who engaged in reflective assessment than for those in the control group, who did not. It also allowed us to see if the students who had lower CTBS scores showed greater gains than students with higher CTBS scores. We found that for the low-CTBS students, the average gain was 25.6 in the reflective assessment classes (an effect size of 1.3σ) and 6.2 in the control classes (an effect size of 0.3σ), a significant difference ($p = 0.02$). In contrast, for the high-CTBS students, the average gain was 18.7 (1.0σ) in the reflective assessment classes and 9.9 (0.5σ) in the control classes, a difference that is only marginally significant ($p = 0.08$). Thus the reflective assessment process, while

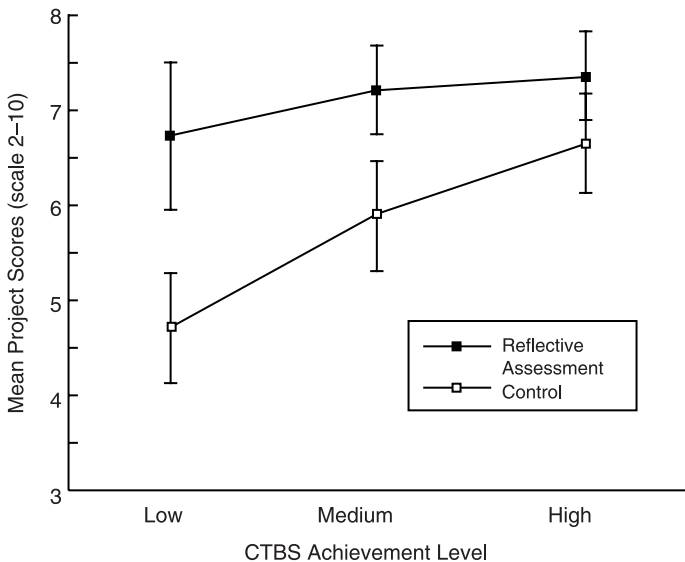


Figure 10.1 Mean scores on their research projects for students in the Reflective Assessment and Control classes, plotted as a function of CTBS achievement level.

beneficial for all students, has the effect of bringing the low-CTBS students up to a level that is closer to that of the high-CTBS students. These results, and those of others (Black & Wiliam, 1998), show that reflection on one's performance is an important part of complex learning, such as mastering scientific inquiry.

Learning Using Explicit Models of Inquiry Meta-Knowledge

Another approach we have taken for developing meta-knowledge and metacognitive practices for inquiry is to have students carry out inquiry projects with the support of a web-based resource called the Web of Inquiry, which provides a model of inquiry goals and fosters metacognitive practices that are intertwined in supporting students' inquiries (Eslinger, 2004; Shimoda, White, & Frederiksen, 2002; White & Frederiksen, 2005a, 2005b; Frederiksen, White, Li, Herrenkohl, & Shimoda, 2008). This resource, and its precursor called "Inquiry Island," provides students with explicit models of inquiry processes, their purposes, and how they are used together in carrying out inquiry. These models are infused with meta-knowledge of science and inquiry.

In addition, we promote metacognitive practices through the use of reflective assessment resources, which students use throughout their research in formatively assessing their progress in accomplishing inquiry goals. These resources include self-assessment tools for formatively evaluating their progress, and a threaded dialogue facility that supports students in asking for and receiving guidance from other students. These assessments are closely aligned with the meta-knowledge that is presented in the models of inquiry processes found in the Web of Inquiry. Our hypothesis is that students can acquire an understanding of the meta-knowledge of scientific inquiry by explicitly using this knowledge as they undertake, and help each other undertake, research projects. In the following, we illustrate how meta-knowledge has been presented in the Web of Inquiry.

The Inquiry Cycle

The Web of Inquiry learning environment includes a top-level structure for inquiry, called the Inquiry Cycle, a research notebook, which scaffolds the goals for each inquiry step, and a community of software "advisors" who "live" on "Inquiry Island." The Inquiry Cycle (see Figure 10.2) depicts the meta-knowledge of inquiry as a cyclical journey from Questioning and Theorizing, through Hypothesizing, to Investigating, and from there to Analyzing Data, Synthesizing a new "Current Best Theory" and providing evidence to support it, and Extending what has been learned. Thus, the inquiry process cycles between a plane of ideas and theories, and a plane of investigations and observations.

Students begin with curiosity about some phenomena and use their current theories and ideas to develop researchable questions and competing hypotheses to investigate. Using these as a guide, they develop plans for an investigation that will provide data, which will allow them to test their hypotheses. Through analyses of their data, they synthesize their theories and findings to come up with their "Current Best Theory" and provide evidence to support it. They then consider ways in which they might extend their theory and suggest further research that needs to be done.

Inquiry Processes, Goals, and Activities

The Inquiry Cycle involves a set of six steps representing processes that are abstract, and which derive their meaning from sets of related activities or subtasks. Each subtask serves as an important goal of that inquiry process. For instance, the inquiry process called "Analyzing Our Data" has three goals for students: (a) to organize your data, summarize

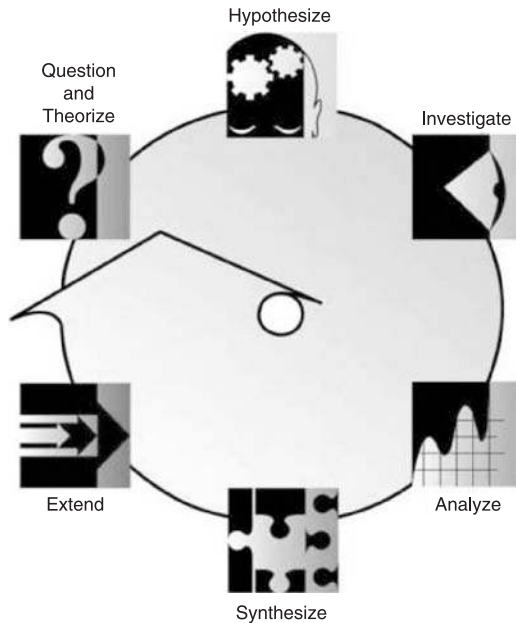


Figure 10.2 The Inquiry Cycle.

it, and create graphs or diagrams to reveal its patterns; (b) to identify all of the patterns in your data and consider their meaning and generality; and (c) to figure out which of your hypotheses are supported by their data and which are not. Each of these goals is the basis for an inquiry subtask that students carry out. The goals therefore link each inquiry process to a set of coordinated activities.

Software Advisors

To provide more detailed meta-knowledge for each inquiry process, the Web of Inquiry includes as resources a set of Inquiry-Task Advisors, one for each of the steps of the Inquiry Cycle. These advisors house knowledge of goals for that inquiry step, reasons why those goals are important, plans and strategies for accomplishing them, and criteria for monitoring their effectiveness. To make it easier for students to keep the advisors straight, each has a name, such as Sydney Synthesizer and Ivy Investigator. Sydney Synthesizer, for instance, can explain what a scientific model is: “Scientific models give you a way to express your ideas. For example, you could make a model to show how you think a spaceship moves through space, or to show the different types of arguments people have, or to show the factors that affect people’s happiness, or to show how the human heart works. This can be done using words, pictures, graphs or diagrams to illustrate things.” He also has information about different kinds of models, including causal models, process models, stage models, structural models, and others, with examples of each for students to study. In addition, he can explain the goals for making models: a model “answers a research question,” “fits findings better than other models,” “seems reasonable,” and “is widely useful.”

The community of advisors also includes general-purpose cognitive, social, and meta-cognitive advisors. For instance Pablo Planner, a metacognitive advisor, has goals for creating a good plan such as “it outlines the goals to be pursued.” Taken together, the community of advisors portrays inquiry as a distributed process in which various

capabilities, like analysis and synthesis, come into play to meet the needs of the research. We personify the cognitive models of each of these processes in the form of advisors so that they can be thought of as roles students may take, or different hats to wear, at different times in their research. The idea is that students will learn to switch in and out of these different hats or roles as they engage in the complex process of scientific inquiry.

Reflection

Reflection involves looking back and evaluating both the process and the products of scientific inquiry. The Web of Inquiry assessment environment provides tools for students to use in reflecting on their own and others' work as they carry out scientific inquiry projects. The assessments in the Web of Inquiry are called "goal assessments" because they are linked to the advisors' goals for each inquiry process. The idea is for students to make self-assessments of their progress as they work, using "developmental" rubrics that give them guidance about how to proceed from one stage of accomplishment to the next. Some examples of goal assessment are shown in Figure 10.3.

The goal assessments include goals that encourage students to bring into play prior inquiry processes they have pursued, such as Theorizing and Analyzing, as appropriate. For example, in the Investigate step, they are asked whether their research design will enable them to test their competing hypotheses. In the Analyze step, they are asked whether they have tested each of their hypotheses. And in the Synthesize step, they are asked to construct a theory that generates predicted outcomes that cover their findings, and this can include findings that may not have been specifically hypothesized.

To foster use of such reflection in social practices of inquiry, we also include a discussion tool, which allows students to ask other groups to review their work and provide feedback to them in response to questions they have asked. Embedding reflective assessment into the inquiry process underscores for students that inquiry is a process of theory formation, testing, and revision, and that these processes are supported by getting feedback from others through collaborative discussions. The goal is to foster an intellectual climate in which students are valued as theory builders, and one in which knowledge is understood to be actively and collectively built, and always open to modification.

<p>Seems Convincing Do your current best theory's descriptions, predictions, and explanations seem believable and fit with everything else you know about your research topic?</p> <ul style="list-style-type: none"> ○ Not plausible This theory doesn't seem believable. It doesn't fit with some things we know about the topic. ○ Fairly plausible This theory seems pretty believable, and it's consistent with most of the things we know about the topic. ○ Highly plausible This theory seems very believable and is consistent with what is known about this topic. 	<p>Fits Findings Does your current best theory generate descriptions, predictions, and explanations that fit all of your research results?</p> <ul style="list-style-type: none"> ○ Doesn't match This theory doesn't fit our research results very well. Some of its descriptions, predictions, or explanations are different from what we found. ○ Partial match This theory is consistent with some of our research results, but it leaves some of them out. ○ Good match This theory fits our research results well. Its predictions and explanations match what we found, and it covers everything that we discovered.
---	--

Figure 10.3 Two examples of reflective assessments students can use while they are working on their Current Best Theory using the Web of Inquiry.

Managing Inquiry

Learning inquiry is complex, as it consists of a large repertoire of inquiry capabilities organized under linked inquiry processes, each with its underlying goals, strategies, and assessments. All of these are in play at various times in carrying out inquiry. If students are to be able to manage their own inquiry, they need to understand when, why, and how these capabilities are engaged in the course of their work. We have referred to this as “metacognitive knowledge for action,” that is, knowledge of how one organizes and manages one’s cognitive and metacognitive processes in the course of their application (White & Frederiksen, 2005b). This involves using meta-knowledge of the processes of inquiry in controlling cognition.

The Web of Inquiry learning environment supports a number of approaches for developing metacognitive knowledge for action. First, aspects of planning and monitoring are represented in the structure of the system. For example, the Inquiry Cycle orders high-level inquiry processes and their sub-goals, so that students learn the logical purpose of the steps within the inquiry process and the value of keeping track of where they are within the Inquiry Cycle. Second, the Web of Inquiry includes metacognitive advisors who provide explicit cognitive models of planning, monitoring, revising, and reflecting, which students can use to guide their work. Third, as a further support for monitoring and reflection, the software provides formative assessments for each inquiry task. These enable students to monitor their progress in achieving the goals associated with each inquiry step as they work. The levels in the assessments are written to represent a developmental sequence and provide ideas for what the students should be thinking about to improve their work on that goal. Fourth, the inquiry tasks typically need the products created in the previous tasks to be available as a condition for that task to proceed. For example, the Investigator needs alternative hypotheses in order to proceed with designing an experiment. If a new task is attempted prematurely, the deficiencies in products of precursor tasks will become apparent, leading students to go back and do additional work on them so that they are good enough to proceed. These four approaches to controlling the flow of inquiry processes are based on a model incorporating distributed responsibility associated with different inquiry roles, coupled with a top-level inquiry goal structure and general habits of monitoring and evaluating one’s progress in relation to goals.

In several studies, we have tested this approach to developing students’ knowledge and use of scientific inquiry. Our most recent and largest study was designed to evaluate how the Web of Inquiry supports students in learning how to carry out scientific inquiry, and teachers in learning how to accurately assess students’ learning (Frederiksen et al., 2008). In this study, the participating teachers agreed to incorporate inquiry projects within their existing curriculum. The seven schools in which our participating teachers taught were chosen to represent a diverse population of students. In doing their inquiry projects, students worked in small groups, usually of two or three, to develop competing theories about some phenomena of interest to them, and then to design and carry out experiments to see which of their theories has the most empirical support.

As they did their research, the students carried out self-assessments and would also work with another group to provide feedback and have conversations about their work. The teachers chose project topics that were appropriate for their particular curriculum. Content areas of the projects included earth and ecological science, physical and materials science, biology, behavioral science, and health and nutrition. Most of the teachers included two projects in their class over a school year, but a few included only one. A total of 271 students’ projects were completed and scored, generally by two scorers. Scores were given for each of the numerous inquiry goals and for nine National Science Education Standards (National Research Council, 1996). The standards were rated on a

four-level scale using the categories (1) “Below Basic,” (2) “Basic,” (3) “Proficient,” and (4) “Advanced.” The mean scoring reliability (G-coefficient for two scorers) is 0.80 for the standards and 0.63 for goals.

To study changes in individual students’ performance brought about by engaging in Web of Inquiry projects, we analyzed the qualities of the students’ initial Web of Inquiry projects, and the extent of improvement in students’ performance between their first and second inquiry projects. Our idea was to identify, based on standards ratings, which aspects of the inquiry process are initially the most well developed, which are most in need of improvement, and how they improved when using the Web of Inquiry environment in carrying out their research.

Mean scores for the students’ first projects for each of the standards are given in Table 10.1. For the most general of the standards, Designing and Conducting a Scientific Investigation, the mean rating on the four-point rating scale is 2.01, which corresponds to the “Basic” level. Only 25% of the projects were rated as “Proficient,” and none were rated as “Advanced.” Thus there was considerable room for improvement after they completed their first project. The mean ratings for three other standards are above this level, while the means for the remaining standards are all below the basic level of 2.0.

The weakest performance on the inquiry standards involves those that pertain to developing alternative theories and hypotheses that can be tested in an investigation, critically analyzing data to distinguish how well the evidence supports each theory, and considering how alternative theories may explain one’s findings, all the while maintaining a skeptical attitude. Students also have difficulty in thinking through ways in which they could extend their work, which is essential if they are to regard inquiry as an ongoing process for improving their knowledge over time in their study of science. These are the areas of inquiry knowledge that are least developed in our sample of classrooms when students carry out a single inquiry project.

To evaluate students’ learning through carrying out Web of Inquiry projects, we analyzed the changes that occurred in students’ inquiry project scores from their first to their second project for four teachers who included two projects in their classes. We found that there is a significant improvement in scores for the overall standard of Designing and

Table 10.1 Mean Performance Levels for the Science Standards

<i>Inquiry standard</i>	<i>Mean</i>
<i>Overall—Designing and conducting a scientific investigation</i>	2.01
<i>Using tools and methods—Using research methods and tools to gather, analyze, and interpret data</i>	2.34
<i>Communicating science inquiry—Communicating scientific procedures and explanations</i>	2.31
<i>Asking scientific questions—Asking questions that can be answered through scientific investigations</i>	2.32
<i>Developing explanations and models—Developing descriptions, explanations, predictions, and models using evidence</i>	1.95
<i>Relating evidence and explanations—Thinking critically and logically to make the relationship between evidence and explanations</i>	1.89
<i>Analyzing alternative explanations—Recognizing and analyzing alternative explanations and predictions</i>	1.76
<i>Showing scientific skepticism—Science advances through legitimate skepticism</i>	1.79
<i>Extending the research—Generating new ideas and methods for extending your research</i>	1.88

Conducting a Scientific Investigation with an effect size of 0.57σ . This effect size should be taken as an underestimate, because it does not reflect students' learning while they are engaged in their first inquiry project. For the first project, 47% of the projects from these four teachers' classes were judged to be proficient and none were judged to be advanced, while for the second project, 70% were judged to be proficient or advanced. This substantial improvement may be attributed to students' working with the Web of Inquiry and also to other aspects of teaching that occurred when the teachers incorporated the Web of Inquiry in their classrooms.

We also determined which standards are most amenable to learning, and which are more difficult for students. We found that there are improvements in project scores for six of the standards: (1) Asking scientific questions, (2) Developing descriptions, explanations, and models, (3) Analyzing alternative explanations, (4) Showing scientific skepticism, (5) Extending your research, and (6) Communicating science inquiry. One of the hallmarks of the Web of Inquiry is the emphasis on having students use their theories to guide their work as they create research questions and hypotheses, identify meaningful patterns in their data, and consider alternative theories in creating and supporting their "current best theory." These findings suggest that using the resources of the Web of Inquiry fosters the development of these abilities. Significant gains in students' communication scores suggest that they are also gaining mastery over the language and practices of scientific presentations and writing.

In contrast, students showed only small improvements for the standard Relating Evidence and Explanation, which is a standard that was also difficult for them in their first project. Relating Evidence and Explanation has three major components: testing how data patterns they have described support predictions from their theories, being careful to identify all salient and consistent data patterns and not to overlook data patterns that might not fit their theories, and considering other plausible theories for explaining their findings. These skills for analyzing and interpreting data from multiple theoretical perspectives present a particularly difficult problem for teaching and learning, and will need to be addressed. Nonetheless, our overall conclusion is that even young students (as young as the fifth grade) can learn to carry out inquiry projects when they are taught, not by leading them step by step through the inquiry processes, but through developing a higher-level meta-knowledge of the structure, purposes, and goals of inquiry processes.

Learning Through Role Playing

We have described how carrying out scientific inquiry is a complex process in which interrelated processes are carried out, and not necessarily in a linear fashion. The decisions about when to move to a different inquiry process or sub-task depend upon active analysis of progress and the readiness of products for supporting future inquiry processes. Such active management depends upon having metacognitive capabilities like planning, monitoring, reflection, and revision. While these are represented in Inquiry Island (and in its web version, the Web of Inquiry) as general-purpose metacognitive advisors, these advisors, unlike the Inquiry-Task advisors, are not directly linked to work on inquiry goals. While their tie-in with accomplishing inquiry processes is implicit, it is not usually made explicit in the software (one example of an exception is when the Investigator refers you to the Planner to plan your investigation). This led us to study whether students' prior knowledge of such metacognitive processes, and whether providing students with additional ways to employ and internalize these metacognitive processes in their inquiry groups, will lead to greater learning using the Inquiry Island approach. We carried out two studies to investigate these questions.

Effects of Students' Prior Knowledge of Metacognitive Processes on Learning Inquiry

The objective of this first study was to evaluate whether students' prior metacognitive knowledge contributes to their development of knowledge and skills in scientific inquiry. This study was carried out in three eighth-grade science classes in an urban school with a teacher who was familiar with inquiry teaching. To investigate this question, we used a test of metacognitive knowledge we developed which we call "Metacognition the Movie." The test was designed to assess how well students understand the concept of metacognition, even with little formal instruction. In this assessment of metacognitive expertise, students watch a movie and then complete a written test. The movie shows two students doing a research project and, from time to time, one of them turns to the camera and expresses a metacognitive thought or sometimes indicates that she has a thought (which is not expressed). In the written assessment, the students are presented with excerpts from the movie's script and are asked to respond to either a justified multiple-choice question or a free-response question about one of these occasions in the movie. The test questions ask students (a) to generate or select an appropriate metacognitive thought for the actor in the movie, (b) to characterize the type of metacognitive thought (planning, checking understanding, and so forth) expressed by the actor, or (c) to explain why that type of metacognition is important at that point in the actors' dialogue.

Different versions of this assessment were given to students as pre-tests and post-tests, at the beginning and end of the school year. The versions were counterbalanced for the two occasions. Scoring was carried out using scoring rubrics that provide descriptions of three response levels. Level 3 responses give metacognitive interpretations of the students' thinking in the movie, level 2 responses describe aspects that are related to the specific work of the students in the movie but do not exhibit metacognition, and level 1 responses are not reflections on the work issues facing the students in the movie. The inter-rater scoring reliability of scoring judgments ranged from 0.82 to 0.88 for the different items.

To test whether students' prior metacognitive knowledge contributes to students' development of knowledge and skills in scientific inquiry, we compared the effects of inquiry instruction for students with low and high metacognition test scores. Our measure of inquiry knowledge is the total score on the Inquiry Test, which was described earlier. For students with low metacognition scores (total scores below the median), there was an increase in mean scores from the pre-test to the post-test with an effect size of 0.30σ , but this increase was not significant ($p = 0.09$). Carrying out the same comparison for students with high metacognition scores (scores above the median), we found there was a significant increase in mean scores ($p < 0.001$), with an effect size of 0.55σ . These results were the same when we controlled for the students' CTBS scores. Thus there is evidence that students with higher pre-test metacognitive capabilities show greater gains in learning inquiry within the curriculum than do students with lower metacognitive capabilities. In other words, students' level of metacognitive expertise predicts who will gain the most from our inquiry curriculum.

We also tested to see if the students' scores on the metacognition test changed as a result of participating in the inquiry curriculum using Inquiry Island. In this case, we found that there were no significant increases in mean scores on the Metacognition the Movie test. This finding led us to the question, "How do we develop *all* students' metacognitive abilities so that they would be better equipped to learn using a resource such as Inquiry Island (or the Web of Inquiry)?"

The Metacognition the Movie test asks students to explain, in the context of collaborative inquiry situations shown in the movie, how metacognition is playing a role in inquiry

and to provide relevant metacognitive thoughts for particular situations that are shown. In the students' work with Inquiry Island, metacognition is practiced primarily through self-assessment using the goal assessments contained in the Inquiry Island software. We hypothesized that adding additional activities in which students engage in collaborative metacognition in their research groups, such as taking on the roles of the various cognitive, social, and metacognitive advisors, will enhance the impact of the inquiry curriculum on students' developing metacognitive capabilities, making them more broadly applicable. We investigated this hypothesis in a second study.

Providing Additional Experiences for Students in Using Metacognition Through Role Playing

To foster students' collaborative inquiry and regulation, we created additional curricular activities in which we used role playing as a way for students to learn the metacognitive skills of planning, monitoring, reflecting, and revising, while practicing them in their group work. We then implemented these ideas in a fifth-grade classroom in an urban public school that was also using Inquiry Island in teaching science through inquiry (White & Frederiksen, 2005a, 2005b).

The students in this class carried out two inquiry projects, and in between doing these projects, they engaged in the role-playing unit in their literacy groups. The purpose of the literacy groups was to understand a fictionalized biography through group discussion. Their discussion was structured as a collaborative inquiry task (with five steps: question, theory, evidence, synthesis, and application). This task was designed to be similar to steps in the Inquiry Cycle they were using in science. Finally, for their second inquiry project, the students carried out research on roles in which they created competing hypotheses about how role playing impacts a group's performance and functioning, and then designed and carried out investigations to see which of their hypotheses were the most accurate. Again they used the Inquiry Island software to guide them through these research projects.

The roles the students played in their literacy groups included cognitive roles (Theory, Evidence, Synthesis, and Application Managers), social roles (Collaboration, Communication, Mediation, and Equity Managers), and metacognitive roles (Planning, Productivity, Reflection, and Revision Managers). The roles are coordinated with the cognitive, social, and metacognitive advisors in Inquiry Island. In their reading groups, students started by playing the cognitive roles, went on to focus on the social roles, and finally played the metacognitive roles.

One-page guides for each role provided students with descriptions of the goals and problems associated with their role, the strategies for achieving each goal and avoiding problems, and "things to say" to implement each strategy (see Figure 10.4 for a sample guide). The students kept journals in which they were asked to plan, monitor, and reflect on the playing of each of their assigned roles. In these journals, and in the curriculum in general, playing a role successfully is defined as getting your group to achieve your role's goals.

Our findings reveal that the students engaged in substantial amounts of role playing in their groups, and that each student became capable of playing multiple roles in a given session. Students not only learned to use the roles they were assigned (or had chosen) to play, they also began to play roles they had never been assigned, which they had seen other students play. This shows that students can *transfer* the knowledge and capabilities they have gained for one role to playing new roles. As the curriculum progressed, the students' role playing became less verbatim (sticking to the guides in deciding what they would say) and more adapted (using their own words, which fit the context well). This suggests that

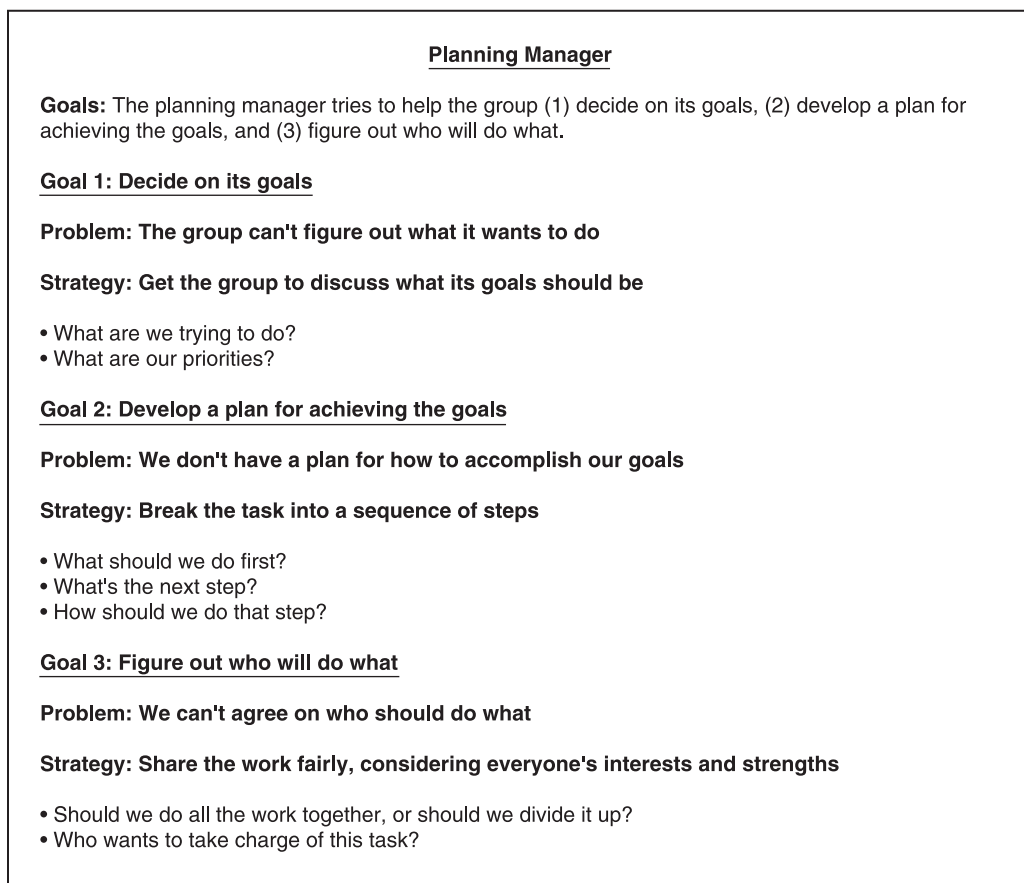


Figure 10.4 The guide for the role of Planning Manager.

they were internalizing the roles and were able to *adapt* them to new situations. When students first undertook each set of roles (social, cognitive, or metacognitive), there was relatively more talk about the goals and strategies for those roles, but as they become more experienced with them, problems and “things to say” increased in relative frequency, with “things to say” being by far the most frequently used component of the guides. This indicates a shift from talking about how to play a role to actually playing the role.

In tests of role understanding and interviews, the students were able to talk about the goals and purposes of each role, about how the different types of roles could be used in new contexts, and about why they would be useful. Their journal entries show a significant increase ($p = 0.004$), over the course of the curriculum, in their understanding of how to get the group to achieve a role's goals. Taken together, these findings suggest that introducing activities in which students play managerial roles can develop young learners' understanding of cognitive, social, and metacognitive capabilities. Students also appreciated the usefulness of the roles in monitoring a group's functioning and in intervening to improve it.

When our inquiry curriculum was expanded to include this type of role playing, we found that there was a significant increase in students' Metacognition the Movie test scores. Fifth-grade students in the school were given this assessment of metacognition at the beginning and end of the school year. The students who undertook this sequence of

pedagogical activities, including Role Playing and the Research on Roles, showed significant gains on this assessment of metacognition ($p = 0.01$) with an effect size of $\sigma = 1.1$, whereas fifth-graders in the school who did not participate did not show a significant gain ($p = 0.37$) with an effect size of $\sigma = 0.3$. The metacognitive behaviors depicted in *Metacognition the Movie* are related to the metacognitive roles the students played in the classroom, so this provides an assessment of whether students can transfer their understanding of metacognition to this new context.

Our findings indicate that students are able to write an appropriate metacognitive remark for a character in the movie, to characterize the type of metacognitive thought expressed by the actor in the movie, and to explain why the metacognition remark is important at that point in the actors' inquiry project. This finding, coupled with the students having demonstrated an ability to pick up roles from others and to adapt their roles to new situations, supports the conclusion that the metacognitive knowledge students have developed through role playing is generalizable. Remember that we did not find any improvements in scores on the Metacognition Test for students who learned inquiry by using *Inquiry Island* without the role playing. Enacting cognitive models of metacognitive capabilities through students' role playing in groups engaged in collaborative inquiry, and then having them do research on the utility of the roles, appears to increase students' understanding of metacognition and its purpose.

Learning Through Inquiry About Inquiry

Perhaps our most important pedagogical goal in developing inquiry and metacognitive expertise is to transform students and classroom environments into self-aware, self-improving systems. We want students to become learners who create theories of their own practices—about what they are doing and why—as they engage in cycles of planning, monitoring, reflecting, and improving. There are a number of ways to enable students to do this, and they all involve an interesting metacognitive move: the recursive application of their inquiry and metacognitive processes in order to improve them.

When we work with students, we call this “inquiry about inquiry” or “research on learning.” Research on inquiry is often carried out around another inquiry that is going on in a group. In other words, two inquiries are going on at the same time: inquiry in the domain of study, and inquiry about how you develop knowledge through inquiry.

Inquiry about inquiry can take a number of forms, which can range from simply reflecting on weaknesses and thinking about how to improve, all the way to fully-fledged scientific inquiry, in which, for example, one invents new strategies and carries out controlled comparisons to determine which are the most effective. For instance, students may focus on a particular advisor's strategies, such as those of Molly Monitor, and evaluate them to see if they are helpful. If they are not, they can then modify them, perhaps by adding some new strategies to make the advice better. They could then test the new strategies as they do their next project, or have other groups try out the new strategies and study how well the strategies work for them. In doing experiments with advisors, they would be developing hypotheses, designing an investigation (thinking of data they will collect), and interpreting their findings in ways that will enable them to improve the advisor's strategies. Testing the strategies on themselves is an example of reflective learning. Testing the strategies on others is essentially doing educational research.

The instructional idea, associated with this example of inquiry about inquiry, is to have students represent their ideas in the form of cognitive models that are embedded in the advisors. To test these ideas, they follow the advisor's advice, or take on its role in their group, while the group undertakes an inquiry project. Students evaluate the group's

ability to use the advisor's advice in various task situations and its usefulness in doing the tasks. They then use this information to revise the advisor, or to change their ideas for where it is useful. There is a built-in validity check to this process for improvement: if their ideas can't be enacted or aren't understandable, or if they are not functionally effective for the task at hand, students will have evidence that their ideas need to be revised. Thus they are applying their knowledge of inquiry to their own cognitive, social, and metacognitive processes needed for inquiry in order to improve them. White and Frederiksen (2005a, 2005b) provide examples of engaging fifth-graders in such inquiry about inquiry.

This notion of inquiry about inquiry is consistent with Schraw and Moshman's (1995) claim that it would be beneficial for students to develop explicit formal theories of their cognitive processes, as well as with Bandura's (1997) idea of self-experimentation and Scardamalia and Bereiter's (1983) notion of children as co-investigators who work with researchers to investigate cognitive processes (in Scardamalia and Bereiter's work, it was the cognitive processes needed for writing). Advocating for inquiry about inquiry represents an extreme position, arguing that not only will children benefit from developing explicit models of their own cognitive and metacognitive capabilities, but also that they will benefit from conducting theory-based, empirical research in which they test competing hypotheses about these capabilities and how to improve them. Our research indicates that this is possible and beneficial for young students (White & Frederiksen, 2005a, 2005b).

Concluding Thoughts

Due to the inherent complexities of scientific inquiry, students need to acquire meta-level expertise in order to engage successfully in the scientific enterprise. Unfortunately, the knowledge and capabilities needed to understand and regulate the processes associated with effective inquiry are often found lacking in students (Kuhn, Black, Keselman, & Kaplan, 2000). Thus it is imperative that we develop tools and teaching methods that emphasize critical metacognitive knowledge and processes. As we have illustrated, this can be accomplished by (a) defining and modeling these processes by providing software advisors that suggest appropriate goals and strategies to pursue, (b) giving students the opportunity to practice using and controlling these processes as they engage in authentic inquiry, (c) having students monitor and reflect on their performance using goal-related criteria, and (d) enabling students to investigate and refine these practices by conducting inquiry into their own inquiry-related processes and practices.

We have argued throughout this chapter that teaching scientific inquiry, infused with metacognitive knowledge and practices, enhances both science education and students' metacognitive development. Such an infusion of metacognition is crucial to effective science education, not only because it facilitates the learning of scientific inquiry, but also because it:

- 1 Fosters the learning of science and learning how to learn in general.
- 2 Builds an understanding of the nature of science and its methods.
- 3 Develops widely useful cognitive, social, and metacognitive capabilities.
- 4 Improves students' self-regulatory capabilities.
- 5 Fosters students' abilities to work together.
- 6 Helps develop students' theories of mind and community.
- 7 Leads to autonomous learners and learning communities.

Metacognitive practices such as self-explanation, in which students create explicit

conceptual models of scientific phenomena, play a role in the generation, understanding, and evaluation of scientific theories (Chi et al., 1989). Meta-representational expertise (diSessa, 2002, 2004) is needed to understand and use different epistemic forms to embody these theories, as well as to represent data in ways that enable students to evaluate and test their theories. Self-regulatory capabilities foster the learning and management of complex inquiry, which also develops students' self-regulatory capabilities in general. For example, self-assessment practices can encourage students to monitor and reflect on how well they are testing their theories and considering alternatives. As we have shown, such self-assessment practices develop students' understanding of inquiry processes, as well as motivating and guiding their improvement (White & Frederiksen, 1998, 2005a, 2005b). Taken together, such metacognitive practices improve students' self-regulatory capabilities and foster their learning of science and learning how to learn in general.

Being exposed to and internalizing explicit models of the cognitive, social, and metacognitive capabilities needed for collaborative inquiry and reflective learning develops skills such as theorizing, investigating, and reflecting (White & Frederiksen, 2005a, 2005b). These models can be embodied by software advisors, like Ivy Investigator and Ricky Reflector, or they can become roles that students play, like Evidence Manager and Reflection Manager (White & Frederiksen, 2005a, 2005b). Alternatively, they can be embedded in scientists' journals that have been created to illustrate these processes in action (Magnusson & Palincsar, 2004). Acquisition of such cognitive and metacognitive skills creates capable individuals, who can engage successfully in generating theories, evidence and arguments as they employ processes that are useful, not only for scientific inquiry, but for learning, problem solving, and critical thinking in general. In addition, getting students to undertake research projects, in which they investigate alternative models for such inquiry processes and metacognitive practices, not only enhances students' awareness of these capabilities, but also provides a path to their development and improvement.

Students learn how to approximate a scientific community as they work together on research projects, engaging in many of the practices of a scientific community (Brown & Campione, 1996). The intertwining of metacognition and inquiry helps to develop a classroom community that is highly aware of the status of its theories and methods, can employ and manage them productively, and is constantly trying to improve all of these. This not only helps students learn how to collaborate, it also leads to a better understanding of the nature of science and its methods, and to an understanding of mind and community in general. Extended immersion in such a learning environment should result in a classroom community that is composed of capable, autonomous learners, who have greater feelings of self-efficacy (Bereiter & Scardamelia, 1987) and can collaborate effectively (Borge, 2007).

The vision is to create students and classrooms that are self-aware, self-improving systems, who create theories about what they are doing and why, as they constantly engage in cycles of planning, monitoring, reflecting, and improving. This can include science classrooms in which students develop and test theories about themselves and their social systems, as well as about the physical and biological world. Conducting research into their own metacognitive practices, for example, is a route to enabling students to learn about scientific theorizing and inquiry, while also fostering their metacognitive development.

Furthermore, introducing students to the idea that they and their communities can be improved, and teaching them how to do this through scientific inquiry intertwined with metacognition, should lead them to an important realization: You don't have to be "born smart." Instead, students can learn how to become more capable individuals and better functioning communities of learners (Dweck & Leggett, 1988). Such a realization should further increase students' feelings of self-efficacy and motivate them to develop the

expertise needed for collaborative inquiry and reflective learning, which should serve them well throughout their lives.

Acknowledgments

This research was funded by the National Science Foundation (grants MDR-9154433, REC-0087583, and REC-0337753). We are grateful to Art Graesser, Larry Yore, Marcela Borge and Tatiana Miller for their helpful comments on an earlier version of this chapter. In addition, we thank the members of our research team for their contributions to this work, as well as the students and teachers who participated in our studies. The views expressed in this chapter are those of the authors and do not necessarily reflect those of the National Science Foundation.

References

- Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education*, 13(1): 1–12.
- Azevedo, R. & Cromley, J. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, 96(3), 523–535.
- Baird, J. R. (1986). Improving learning through enhanced metacognition: A classroom study. *European Journal of Science Education*, 8(3), 263–282.
- Baird, J. R. & White, R. T. (1996). Metacognitive strategies in the classroom. In D. F. Treagust, R. Duit, & B. J. Fraser (Eds.), *Improving teaching and learning in science and mathematics*. New York: Teacher College Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
- Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–73.
- Blank, L. M. (2000). A metacognitive learning cycle: A better warranty for student understanding. *Science Education*, 84(4), 486–506.
- Blumenfeld, P., Soloway, E., Marx, R., Krajcik, J., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, 26, 369–398.
- Borge, M. (2007). *Regulating social interactions: Developing a functional theory of collaboration*. Doctoral dissertation, University of California at Berkeley.
- Bransford, J. & Schwartz, D. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24(1), 61–100.
- Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 60–108). Hillsdale, NJ: Erlbaum.
- Brown, A. (1988). Motivation to learn and understand: On taking charge of one's own learning. *Cognition & Instruction*, 5(4), 311–321.
- Brown, A. & Campione, J. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289–325). Mahwah NJ: Lawrence Erlbaum Associates.
- Campione, J. C. (1984). Metacognitive components of instructional research with problem learners. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and learning* (pp. 109–132). West Germany: Kuhlhammer.
- Carey, S. & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist*, 28(3), 235–251.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-Explanations:

- How students study and use examples in learning to solve problems. *Cognitive Science*, 13(2), 145–182.
- Chinn, C. A. & Malhotra, B. A. (2002). Epistemologically authentic reasoning in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86, 175–218.
- Collins, A. & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational Psychologist*, 28(1), 25–42.
- Corno, L. (1986). The metacognitive control components of self-regulated learning. *Contemporary Educational Psychology*, 11, 333–346.
- Davis, E. (2003). Prompting middle school science students for productive reflection: generic and directed prompts. *The Journal of the Learning Sciences*, 12(1), 91–142.
- Dewey, J. (1910). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: D. C. Heath.
- diSessa, A. A. (2002). Students' criteria for representational adequacy. In K. Gravemeijer, R. Lehrer, B. van Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 105–129). Dordrecht: Kluwer.
- diSessa, A. A. (2004). Metarepresentation: Naïve competence and targets for instruction. *Cognition and Instruction*, 22(3), 293–331.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287–313.
- Duschl, R. (2007). Quality argumentation and epistemic criteria. In M. Jiménez-Aleixandre & S. Erduran (Eds.), *Argumentation in science education: Perspectives from classroom-based research*. Springer Netherlands.
- Duschl, R. & Osborne, J. (2002). Argumentation and discourse processes in science education. *Studies in Science Education*, 38, 39–72.
- Dweck, C. & Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256–273.
- Eslinger, E. (2004). *Student self-assessment in an interactive learning environment: Technological tools for scaffolding and understanding self-assessment practices*. Doctoral dissertation, University of California at Berkeley.
- Frederiksen, J. & White, B. (1997). Cognitive facilitation: A method for promoting reflective collaboration. In *Proceedings of the Second International Conference on Computer Support for Collaborative Learning*. Mahwah, NJ: Erlbaum.
- Frederiksen, J. R. & White, B. Y. (2002). Conceptualizing and constructing linked models: Creating coherence in complex knowledge systems. In P. Brna, M. Baker, K. Stenning, & A. Tiberghien (Eds.), *The role of communication in learning to model* (pp. 69–96). Mahwah, NJ: Lawrence Erlbaum Associates.
- Frederiksen, J. R., & White, B. Y., & Gutwill, J. (1999). Dynamic mental models in learning science: The importance of constructing derivational linkages among models. *Journal of Research in Science Teaching*, 36(7), 806–836.
- Frederiksen, J. R., White, B. Y., Li, M., Herrenkohl, L. R., & Shimoda, T. (2008). *Classroom formative assessment: Investigating models for evaluating the learning of scientific inquiry*. Final report to the National Science Foundation. University of Washington: available from the authors.
- Georghiades, P. (2004). From the general to the situated: three decades of metacognition. *International Journal of Science Education*, 26(3), 365–383.
- Giere, R. (1991). *Understanding scientific reasoning*, 3rd ed. Fort Worth: Harcourt Brace Jovanovich.
- Gilbert, S. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*, 28(1), 73–79.
- Graesser, A., McNamara, D., & VanLehn, K. A. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist*, 40(4), 225–234.
- Halloun, I. A. (2004). *Modeling theory in science education*. Netherlands: Springer.
- Herrenkohl, L. R., Palincsar, A., Dewater, L., & Kawasaki, K. (1999) Developing scientific communities in classrooms: A sociocognitive approach. *Journal of the Learning Sciences*, 8(3&4) 451–493.

- Hestenes, D. (1987). Toward a modeling theory of physics instruction. *American Journal of Physics*, 55, 440–454.
- Hogan, K. (1999). Thinking aloud together: A test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching*, 36(10), 1085–1109.
- Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, 77(3), 319–337.
- Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, 18(4), 495–523.
- Kuhn, D. & Pearsall, S. (1998). Relations between metastrategic knowledge and strategic performance. *Cognitive Development*, 13, 227–247.
- Lehrer, R. & Schauble, L. (2000). Modeling in mathematics and science. In R. Glaser (Ed.), *Advances in instructional psychology: Volume 5. Educational design and cognitive science* (pp. 101–159). New Jersey: Lawrence Erlbaum.
- Lehrer, R. & Schauble, L. (2005). Developing modeling and argument in elementary grades. In T. A. Romberg, T. P. Carpenter, & F. Dremock (Eds.), *Understanding mathematics and science matters*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Linn, M. C., Davis, E. A., & Bell, P. (Eds.). (2004). *Internet environments for science education*. Mahwah, NJ, Lawrence Erlbaum Associates.
- Loh, B., Radinsky, J., Reiser, B. J., Gomez, L. M., Edelson, D. C., & Russell, E. (1997). The Progress Portfolio: Promoting reflective inquiry in complex investigation environments. In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Proceedings of Computer Supported Collaborative Learning '97*. Ontario, Canada.
- Magnusson, S. J. & Palincsar, A. (2004). Learning from text designed to model scientific thinking in inquiry-based instruction. In E. Saul (Ed.), *Crossing borders in literacy and science instruction* (pp. 316–339). Newark, DE: International Reading Association.
- Mellar, H., Bliss, J., Boohan, R., Ogborn, J., & Tompsett, C. (Eds.). (1994). *Learning with artificial worlds: Computer based modeling in the curriculum*. Washington, DC: The Falmer Press.
- Metz, K. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cognition and Instruction*, 22(2), 219–290.
- Metz, K. E. (2000). Young children's inquiry in biology: Building the knowledge bases to empower independent inquiry. In J. Minstrell & E. H. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 371–404). Washington, DC: American Association for the Advancement of Science.
- National Research Council (1996). *The national science education standards*. Washington, DC: National Academy Press.
- National Research Council (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academy Press.
- Perkins, D. N. & Grotzer, T. A. (2005). Dimensions of causal understanding: The role of complex causal models in students' understanding of science. *Studies in Science Education*, 41(1), 117–165.
- Pintrich, P. & De Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40.
- Popper, K. (1963) *Conjectures and refutations*. London: Routledge and Kegan Paul.
- Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist*, 40(4), 235–244.
- Sandoval, W. A. & Reiser, B. J. (2004). Explanation driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, 88(3), 345–372.
- Scardamalia, M. & Bereiter, C. (1983). Child as coinvestigator: Helping children gain insight into their own mental processes. In S. Paris, G. Olson, & H. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 61–82). Hillsdale, NJ: Erlbaum.
- Scardamalia, M. & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *The Journal of the Learning Sciences*, 1(1), 37–68.

- Scardamalia, M. & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, 3(3), 265–283.
- Schauble, L., Glaser, R., Duschl, R., Schultz, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *The Journal of the Learning Sciences*, 4(2), 131–166.
- Schoenfeld, A. (1983). Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. *Cognitive Science*, 7, 329–363.
- Schoenfeld, A. H. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 189–215). Hillsdale, NJ: Lawrence Erlbaum.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.
- Schraw, G., Crippen, K. J., & Hartley, K. D. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, 36(1–2), 111–139.
- Schraw, G. & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7, 351–371.
- Schunk, D. & Schwartz, C. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18(3), 337–354.
- Schunk, D. & Zimmerman, B. (Eds.). (1998). *Self-regulated learning: From teaching to self-reflective practice*. New York: Guilford.
- Schwartz, D., Chase, C., Chin, D., Opezzo, M., Kwong, H., Okita, S., Biswas, G., Roscoe, R., Jeong, H., & Wagster, J. (2009). Interactive metacognition: Monitoring and regulating a teachable agent. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 340–358). New York and London: Routledge.
- Schwarz, C. & White, B. (2005). Meta-modeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, 23(2), 165–205.
- Shimoda, T., White, B., & Frederiksen, J. (2002). Student reflective inquiry: Increasing levels of agency through modifiable software advisors. *Science Education*, 86, 244–263.
- Slotta, J. and Chi, M. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction*, 24(2), 261–289.
- Smith, C., Maclin, D., Houghton, C. and Hennessey, M. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction*, 18(3), 349–422.
- Smith, C., Snir, J., & Grosslight, L. (1992). Using conceptual models to facilitate conceptual change: The case of weight-density differentiation. *Cognition and Instruction*, 9, 221–83.
- Smith, C. & Wenk, L. (2006). Relations among three aspects of first-year college students' epistemologies of science. *Journal of Research in Science Teaching*, 43(8), 747–785.
- Stewart, J., Cartier, J. L., & Passmore, C. M. (2005). Developing understanding through model-based inquiry. In M. S. Donovan & J. D. Bransford (Eds.), *How students learn* (pp. 515–565). Washington, DC: National Research Council.
- Toth, E., Suthers, D., and Lesgold, A. (2002). Mapping to know: The effects of evidence maps and reflective assessment on scientific inquiry skills. *Science Education* 86(2), 264–286.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. and Trans.). Cambridge, England: Cambridge University Press.
- White, B. Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction*, 10(1), 1–100.
- White, B. Y. & Frederiksen, J. R. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*, 42, 99–157.
- White, B. Y. & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3–118.
- White, B. Y. & Frederiksen, J. (2005a). *Modeling, developing, and assessing scientific inquiry skills using a computer-based inquiry support environment*. Final report to the National Science Foundation. UC Berkeley: available from the first author.

- White, B. Y. & Frederiksen, J. R. (2005b). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist*, 40(4), 211–223.
- White, R. & Gunstone, R. (1989). Metalearning and conceptual change. *International Journal of Science Education*, 11, 577–586.
- Wilensky, U. & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology*, 8(1), 3–19.
- Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, 92(5), 941–967.
- Winne, P. & Hadwin, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Erlbaum.
- Yore, L., Pimm, D., & Tuan, H. (2007). The literacy component of mathematical and scientific literacy. *International Journal of Science and Mathematics Education*, 5(4), 559–589.
- Zimmerman, B. (1989). A social cognitive view of self-regulated academic learning. *Journal of Experimental Psychology*, 81(3), 329–339.
- Zimmerman, B. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In D. Schunk & B. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice*. New York: Guilford.
- Zimmerman, B. & Schunk, D. (Eds.). (2001). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). Mahwah, NJ: Erlbaum.
- Zohar, A. & Dori, Y. (2003). Higher-order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences*, 12(2), 145–181.

11 The Enigma of Mathematical Learning Disabilities

Metacognition or STICORDI, That's the Question

Annemie Desoete

Department of Experimental Clinical and Health Psychology, Ghent University & Arteveldehogeschool, Ghent, Belgium

Mathematical learning disabilities (MLD) are common developmental disorders in childhood. The term refers to a significant degree of impairment in a child's mathematical skills. First, the problems are not a part of normal development, and remediation does not lead to improvements. Second, the problems are not originated in mental retardation or impairments in general intelligence. The child's level of attainment is substantially below that expected for a child of the same mental age. Furthermore, the difficulties are part of a developmental trajectory. Third, problem onset must be early in the development of the child. In Flanders, most learning disabilities are not detected until primary school, when reading disabilities often become obvious. Likewise, many children with MLD in primary school do not encounter severe difficulties with preliminary mathematics learning in kindergarten, although in some cases inefficient counting and seriation was already present as a "marker" (Grégoire, 2005; Stock, Desoete, & Roeyers, 2007). Nevertheless, most children with MLD in Belgium are detected in grade 1 when they have to master addition and subtraction and, in some cases, even later (in grade 3), when they have to learn to retrieve quickly the times tables or to select and apply various problem-solving strategies in addition to the basic mathematical operations. Fourth, the impairments cannot be attributed to external factors that could provide sufficient evidence for academic failure. The fifth and last criterion states that the developmental disorders of academic skills are not directly due to uncorrected visual or hearing impairments (Desoete, 2008; Desoete, Roeyers, & De Clercq, 2004; Stock, Desoete, & Roeyers, 2006).

The prevalence of MLD in children and adolescents will vary depending on the criteria used to define the "substantially below" performances (Barbarese, Katusic, Colligan, Weaver, & Jacobsen, 2005). Geary (2004) revealed that between 5% and 8% of school-age children have some form of mathematical disabilities. These figures are confirmed in different countries by several researchers (Dowker, 2004). Although this prevalence of mathematical disabilities seems as high as the prevalence of reading disabilities, the research focus on the domain of this vulnerable group of children and adolescents still remains limited. We agree with Ginsburg (1997) and McNamara (2007) that the problems of children with mathematical disabilities have been underestimated and are rarely heard.

In this chapter, we will address a conceptual model with protective and deficient functions of children and adolescents with MLD. In addition, we will examine whether metacognitive problems in some children with MLD can be considered as a retardation or deficit perspective. Moreover, we focus on the assessment and training of metacognition. Finally, we discuss how STI(mulation), CO(mpensation), R(emediation) and DI(spensation) (STICORDI) devices are needed if metacognition training programs fail.

Model of Analysing Problems with MLD

Reflecting on developmental issues in numeracy and our own experiences in the domain of MLD, we developed a model for young children of mathematical problem solving, with cognitive and metacognitive underpinnings (see Figure 11.1). We tested the model for conceptual accuracy and clinical relevance on different samples of children and adults with average intelligence and MLD. Parts of the model have been used to assess mathematical competencies of children with MLD (Desoete & Grégoire, 2007; Desoete & Roeyers, 2005). According to the model, mathematical problem solving depends on adequate procedural skills, fact retrieval skills, visual spatial skills and number knowledge reflecting in four phenotypes and a combination of these “phenotypes” (PT) in MLD (see Procedural PT, Semantic PT, VisuoSpatial PT and Number Knowledge PT). Landerl, Bevan and Butterworth (2004) advise that we first need to search for the “causes” (see C) of MLD instead of searching for subtypes and phenotypes. There are several models trying to explain MLD. One of these models (C1) originate MLD in subitizing deficits (i.e., deficits in knowing how many items lie within a visual scene) that are present since birth. Those children cannot subitize when they are babies and have difficulties in the comparison of quantities. When they become toddlers, they do not develop a mental number line and have difficulties in counting (Reeye & Reynolds, 2004). Another model (C2) originates MLD in a disturbed mental number line (Ansari et al., 2003; Butterworth, 1999). Some models focus on immature problem-solving strategies (C3; Geary, 2004) and slow serial elaboration (Geary & Brown, 1991), a deficit in the working memory (C4) or in semantic long-term memory (C5; Landerl et al., 2004), problems with visual spatial elaboration (C6; Geary, 2004) and executive deficits (C7; Passolunghi & Siegel, 2004). Further research in these areas is needed, because a good understanding of different causes (C1–7) can help to identify children’s strengths and weaknesses (Grégoire, 1997).

However, the developmental trajectory of children and adults with MLD also depends on their efficient skills and capacities to overcome and cope with problems. Important compensating “modifying processes” (MP) seem to be the intelligence (MP 1), motivation (MP 2), attention (MP 3), and metacognition (MP 4). The impact of poor mathematical skills is even bigger in below-average intelligence children (deficit in MP 1), in children with ADHD (deficit in MP 3), and in children with low motivation (deficit in MP 2) or a

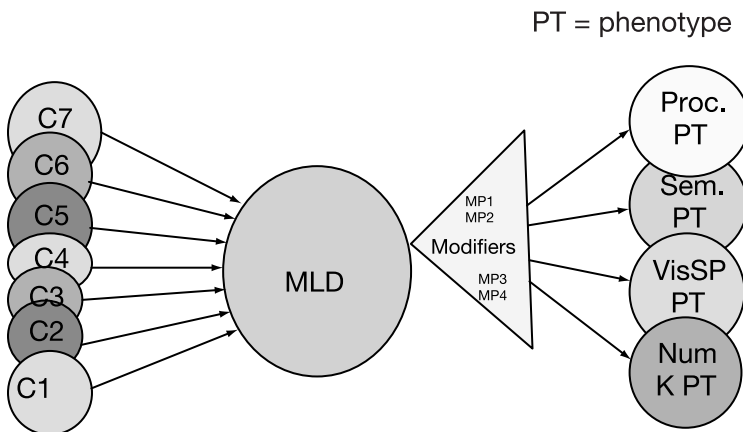


Figure 11.1 Conceptual model on cognition and metacognition in subjects with (MLD).

Note: C1–7 = initial Causal (C) processes; MLD = mathematical learning disability; MP = modifying processes (such as MP1 = intelligence, MP2 = motivation, MP3 = attention, and MP4 = metacognition); PT = phenotype; Proc. PT = Procedural Phenotype of MLD, Sem. PT = Semantic Phenotype of MLD, VisSP PT = VisuoSpatial Phenotype of MLD, and Num K PT = Number Knowledge Phenotype of MLD.

lack of metacognitive knowledge and skills (deficit in MP 4). If children have problems following the sequence in long algorithms (Procedural or Proc PT), they first must be intelligent enough (MP 1) and motivated (MP 2) to plan efficiently (MP 4), and then work even more concentrated (MP 3) and organized (MP 4) on the task than peers would do in a similar situation. Intelligence, attention, motivation and the use of metacognitive knowledge and skills are important for the chances of success of those students. Shalev, Manor, and Gros-Tsur (2005) showed that children with mathematical disabilities in combination with ADHD (deficit in MP 3) or a below-average intelligence (deficit in MP 1) were more vulnerable and profoundly impaired than children with mathematical disabilities alone. A comorbidity with motor problems, such as DCD, is less impairing than MLD and ADHD (Desoete, 2008), because attention (MP 3) is one of the coping strategies or “modifiers,” and motor skills cannot be seen as a modifying process (MP).

Cognitive Underpinnings of MLD

In general, four descriptive cognitive underpinnings of mathematical disabilities are currently found in the scientific research: procedural deficits, semantic memory deficits, visual spatial deficits, and number knowledge deficits.

First, cognitive underpinnings concern difficulties in the execution of arithmetic procedures (see procedural phenotype or Proc PT in Figure 11.1). Children exhibit difficulties in sequencing the multiple steps in complex procedures and make a lot of errors in executing them. These children have a poor understanding of the concepts they use for the procedures (Geary, 2004).

A second pattern or phenotype (PT) of cognitive impairments in MLD is described as problems due to semantic memory deficits (see semantic phenotype or Sem. PT in Figure 11.1). Geary (2004) describes children having difficulties in retrieving mathematical facts. Cornoldi and his colleagues describe children with a lower accuracy of mental calculation, slower speed of mental and written calculation, lower enumeration speed and difficulties with retrieving numerical facts (Cornoldi & Lucangeli, 2004).

A third cognitive phenotype of MLD that is often described in the literature contains a conjunction of visual spatial disabilities in the arithmetic domain (see visual spatial phenotype or Vis.Sp. PT in Figure 11.1). This subtype is characterized by problems with insight in and conceptions of space. Those cognitive deficits are typically translated into difficulties in situating numbers on the number line, shuffling numbers in big figures, and difficulties in the understanding of geometry. Cornoldi and colleagues (2003) described the presence of a discrepancy between verbal and non-verbal IQ scores and the failure in visual spatial memory. Geary (2004) describes the visual spatial subtype as a profile with difficulties in spatially representing numerical and other forms of mathematical information and relationships. Children with this kind of problem also frequently misinterpret spatially represented information.

The fourth cognitive phenotype of MLD involves a less well described cognitive deficit in MLD and concerns the number knowledge deficits (see number knowledge phenotype or Num K PT in Figure 11.1). People with this sort of MLD do not know the specific positions for units, tens, and hundreds. Problems often emerge in number reading, number writing, number production and number knowledge. The existence of these cognitive deficits is confirmed by the findings of Cornoldi and his colleagues (e.g., Cornoldi et al., 2003; Cornoldi & Lucangeli, 2004). In their five-level approach, they also describe a pattern of impairments in numerical knowledge. Those impairments can be due to difficulties in size comparison, word-number transcoding, and enumeration or number dictation. Difficulties in the reading and comprehension of Arabic numbers are also reported by von Aster (2000). These children have difficulties in the comprehension of the Arabic

notational system and in transcoding between the different modalities. Another characteristic cognitive feature is the difficulty in the place value system and vertical alignments (von Aster, 2000).

Most of the procedural, factual retrieval, visual spatial, and number knowledge problems of children and youth with MLD are life-affecting. As adults, these children and youth will need to confront problems that pervade everyday life, such as following directions or using maps, charts, graphs, tables, or plans. Difficulties also remain in situations where adults with MLD have to pay bills, consult bus or train schedules, read instructional leaflets, or determine expiry dates on products. In addition, the impact of poor mathematical skills on employment prospects is even bigger than the influence of poor reading skills (Dowker, 2004). Moreover, it is clear that the different descriptive cognitive underpinnings in mathematical disabilities cannot be seen separately. The onset of one cognitive deficit will influence the functioning and development of other cognitive skills. Difficulties in number knowledge or visual spatial problems, for instance, will slow down calculation procedures, which can have an influence on the memorization of number facts. Besides, because mathematical disabilities are a developmental disorder, manifestations of the disability are related to age and developmental processes. Ginsburg (1997) points out that some children can outgrow one cognitive phenotype into another one. As a matter of fact, the profiles of the children we meet in practice constitute features of different phenotypes described above (for a review see Stock et al., 2006).

Metacognitive Underpinnings of MLD

Presently, it is widely accepted that metacognition (see MP 4) influences mathematical problem solving (e.g., Desoete, Roeyers, Buysse, & De Clercq, 2002; Goos, 2002; Montague, 1998; Veenman, Kok, & Blöte, 2005). Proficient students are assumed to select appropriate strategies and adjust behavior to changing task demands, making use of the awareness of previous knowledge and selecting appropriate study behavior (Montague, 1998).

Metacognition has been found to be instrumental in challenging tasks in mathematics (Tanner & Jones, 2002). Metacognition is involved in nearly every aspect of mathematical problem solving, from the initial stage of building an appropriate representation of the problem to the final stage of interpretation and checking the outcome of the calculations (Verschaffel, 1999). Metacognition prevents “blind calculation” or a superficial “number crunching” approach in mathematics (e.g., answering “43” to the exercise “40 is 3 more than . . .,” because “more” is always translated into additions). Moreover, metacognition allows students to use the acquired knowledge in a flexible, strategic way (Lucangeli, Cornoldi, & Tellarini, 1998). Tobias and Everson (2000) found that metacognitive skills were especially necessary to evaluate what you know or do not know, to monitor learning, and to plan learning new skills.

A substantial amount of data have been accumulated on four metacognitive skills: prediction or goal setting, planning, monitoring or regulation and control or evaluation (e.g., Desoete & Roeyers, 2005; Sperling, Howard, Staley, & Du Bois, 2004; Veenman, Van Hout-Wolters, & Afflerbach, 2006; Zimmerman, 2000). In mathematics, prediction enables children to think about the learning objectives, proper learning characteristics, and the available time. Planning involves analyzing exercises (e.g., “It is a multi-digit exercise in a number problem format”) and retrieving and activation of metacognitive knowledge and skills. Monitoring is related to questions such as “Am I following my plan?”, “Is this plan working?”, “Should I use paper and pencil to solve the exercise?” and so on. Typically, metacognitive monitoring and control would be considered the hub of

self-regulated learning (SRL; see Veenman et al., 2006). Monitoring and assessing the adequacy of specific strategy use may lead a learner to apply a different strategy. In evaluation there is self-judging of the answer and of the process of getting to this answer. The evaluation skill can be defined as the retrospective reflections that take place after an event has transpired (Brown, 1987), whereby children look at what strategies were used and whether or not they led to a desired result. Specifically, children reflect on the outcome and the understanding of the problem and the appropriateness of the plan, the execution of the solution method as well as on the adequacy of the answer within the context of the problem. Evaluation makes children evaluate their performance and compare their task performance with other people so that the final result can be used to locate errors in the solution process (Lucangeli et al., 1998).

From a developmental point of view, metacognitive knowledge precedes metacognitive skills (Flavell, 1976, 1979). Metacognitive self-regulating skills grow until early adulthood (Veenman, Wilhelm, & Beishuizen, 2004) and likely beyond. Metacognition allows students to use acquired knowledge in a flexible, strategic way (Desoete, Roeyers, & Buysse, 2001; Focant, Grégoire, & Desoete, 2006; Lucangeli et al., 1998).

In the literature, several models are distinguished concerning the relationship between cognition and metacognition. The intelligence model states that metacognitive skills are part of intelligence (Sternberg, 1985, 1990). In other words, metacognitive skills are considered to be a manifestation of intelligence, and, consequently, they do not have a separate influence on learning outcomes. The independence model considers intelligence and metacognitive skills as completely independent predictors of study performances: both intelligence and metacognitive skills have their own influence on academic performance. An interesting question related to this independence model is whether or not metacognitive skills could compensate for intelligence (compensation hypothesis), and if training of these skills is possible. A third mixed model states that intelligence and metacognitive skills have some variance in common, but that metacognitive skills do have an additional value on top of intelligence in the prediction of study performances.

Several studies on university students provided evidence in favor of the mixed model for a variety of tasks. Yet, the relationship between mathematics, metacognition and intelligence was found to depend on several other contributory factors. For instance their relationship was found to differ according to the prior knowledge of children. The relationship also differed according to the nature of the metacognitive component that was looked at, the age of the children and a number of instructional variables. Alexander, Carr, and Schwanenflugel (1995) found that highly intelligent children, independent of age, had more declarative metacognitive knowledge than less intelligent peers. However, with regard to metacognitive monitoring, no such differences were found in young children. Furthermore, it was found that a teaching approach that provided structure promoted mathematics proficiency in less intelligent children with few metacognitive skills, but not in less intelligent children with good metacognitive skills. Bright children did well, irrespective of the instructional variables (Cronbach & Snow, 1977). In addition, the relationship was found to differ in children without learning disabilities and in a clinical group of children with learning disabilities (Desoete & Roeyers, 2006).

Metacognitive Retardation or Deficits in Children with MLD

An unresolved issue concerns the question about whether the low mathematics performances in children with MLD can be explained by the delay or deficit hypotheses (Geary, 2004).

According to Wong (1996) the assumption that students with learning disabilities

lack metacognitive skills is invalid. Instead these children appear to have less sophisticated metacognitive skills than peers without learning disabilities. Furthermore, low metacognitive scores in children with learning disabilities are considered by Borkowski and Thorpe (1994) to be the result of insufficient maturity in the development of the regulation of mathematical cognition. In this case metacognitive differences between children with and without learning disabilities can be explained according to the maturation delay or retardation hypothesis.

Another possible explanation is the deficit hypothesis, whereby metacognition is considered a deficit in children with learning disabilities. In the case of the deficit hypothesis, children with learning disabilities would have different or disharmonically developed metacognitive knowledge and skills, not comparable to the skills and knowledge of younger children matched at mathematical performance level. Davidson and Freebody (1986) found the deficit hypothesis incapable of explaining some of their research data.

As suggested by the delay hypothesis, our data show a metacognitive profile (strengths and problems) for children with MLD that is not comparable on all aspects to the profile of other younger children (Desoete & Roeyers, 2006). In addition, not all children with MLD and a minority of the children without MLD appear to show less developed metacognitive skills. Approximately a third of children with MLD have inappropriate or less developed metacognitive skill. More specifically, this is the case for around two-thirds of the MLD children with a combined disability, one-half of the children with procedural MLD (Proc PT), and only around 5% of the children with fact retrieval or semantic memory deficits (sem PT). Furthermore, 59% of the children with MLD and less developed and inappropriate metacognitive skills have problems with prediction and evaluation skills. Most third-grade children with less developed metacognitive skills have problems with assessing the actual level of difficulty of problems. More than half of the third-grade children with at least average mathematics skills and less developed metacognitive skills also have problems with evaluating their task performance. The study suggests that there is a spectrum of MLD, with different cognitive and metacognitive problems and strengths in young children. It is also apparent that the nature of metacognitive delayed or deficient skills of children with or without learning problems differs. Therefore, it is advisable to be cautious when applying findings in the metacognitive literature from children without MLD to children with MLD.

Metacognitive Assessments of Children with MLD

Different methods to assess metacognitive knowledge and skills make study outcomes often difficult to compare (Desoete & Roeyers, 2006).

Prospective methods, such as self-report questionnaires and hypothetical interviews, are frequently used to assess metacognitive skills or beliefs (e.g., Miranda, Garcia, Marco, & Rosel, 2006). Several studies underlined the importance of these instruments (e.g., Busato, Prins, Elshout, & Hamaker, 1998). However Veenman (2005) pointed to the limited explained variance towards learning outcomes of prospective assessment methods.

Retrospective techniques, such as questionnaires, interviews, and stimulated recall situations, are also being used to assess metacognitive skills. Several authors fruitfully used or combined such instruments (e.g., Efklides, 2001). Moderate correlations were demonstrated between prospective and retrospective measures (Veenman, 2005). However, retrospective verbal reports of higher-order skills often seem to lack accuracy (Nisbett & Wilson, 1977) and a very limited explained variance was found for such techniques (Veenman, 2005).

In addition to prospective and retrospective techniques, concurrent assessment, such as

think-aloud protocols (e.g., Artzt & Armour-Thomas, 1992; Azevedo, Greene, & Moos, 2007), systematic observation of regulation and planning behavior (e.g., Pugalee, 2001), eye-tracking studies, error detection studies, etc. can take place (e.g., Metcalfe, 1998). Protocol analyses were found to be very accurate but time-consuming techniques to assess metacognitive skills (e.g., Azevedo & Witherspoon, this volume; Pressley, 2000). Veenman (2005) found a high correlation between protocol analyses and systematic observation measures. Moreover, a well explained variance towards learning performances was found. However, the correlation between concurrent and non-concurrent (prospective or retrospective) measures was very low and an adequate level of verbal fluency seems necessary to avoid interference of the verbalization with the target task (Thorpe & Satterly, 1990).

Multi-method techniques are now being used. Often these techniques combine prospective and concurrent or concurrent and retrospective measures of metacognitive skills and/or knowledge (e.g., Lucangeli & Cabrele, 2006). For example, students are asked before and after the solution of a mathematics task to assess the difficulty of the task, the correctness of the solution (conceived or produced), and the effort required, and to make subjective estimations about the use of problem-solving strategies. In calibration studies (e.g., Winne, 1997; Schraw, 2000) a comparison is made of whether the prediction before the tasks ("calibration" or comprehension paradigm) or the evaluation after a task ("performance calibration" or post-diction paradigm) corresponds with the actual performance on the task. Calibration studies are therefore most related with metacognitive knowledge assessment (see Schraw, this volume).

Tobias and Everson (2000) developed the Metacognitive Knowledge Monitoring Assessment (KMA) that compares what students think (or predict) they know or do not know and what they really know or do not know. This research design is very similar to our own paradigm to measure predictions and evaluations with EPA2000 (Desoete & Roeyers, 2006).

EPA2000 (De Clercq, Desoete, & Roeyers, 2000; Desoete, Roeyers, Buysse, & De Clercq, 2002) has a three-part (metacognitive prediction, mathematical problem solving, metacognitive evaluation) assessment. Children have to predict and evaluate on 80 mathematical problem-solving tasks, including tasks at grade 1, 2, 3, and 4. EPA2000 includes tasks on the comprehension of numbers and operation symbols, number system knowledge, mental arithmetic, procedural arithmetic, and word problems. In the measurement of prediction, children are asked to look at exercises without solving them and to predict, on a 4-point rating scale, whether they will be successful in this task. Children have to evaluate after solving the different mathematical problem solving tasks on the same 4-point rating scale. In EPA2000, children have to comprehend the instruction and to click on the answer with the mouse. With EPA2000, the accuracy in problem solving is scored as well as the accuracy of predictions and evaluations. Children can give four ratings (1 absolutely sure I am wrong, 2 sure I am wrong, 3 sure I am correct, 4 absolutely sure I am correct). Metacognitive predictions or evaluations are awarded two points whenever they correspond to the child's actual performance on the task (predicting or evaluating 1 and doing the exercise wrong and rating 4 and doing the exercise correctly). Predicting and evaluating, rating 1 or 3 receive 1 point whenever they correspond. Other answers do not gain any points, as they are considered to represent a lack of off-line metacognition. For the mathematical problem solving, children obtain 1 point for every correct answer. The three scores (prediction, mathematical problem solving and evaluation) are unrelated. For instance, in theory a child can obtain maximum scores for prediction, a zero score for mathematics and a medium score for evaluation.

Although some researchers question the trustworthiness of teacher questionnaire data, reviews indicate that teachers' judgments can serve as worthy assessments of students'

achievement-related behaviors triangulated with data gathered by other protocols (Winne & Perry, 2000). Furthermore, the teacher's perception of students' use of skills was found to be an important predictor of academic performances in children with learning disabilities (Meltzer, Roditi, Houser, & Perlman, 1998). In addition, teacher questionnaires were found to have some value added in the evaluation of metacognitive skills. A recent study revealed that metacognitive skills assessed by teacher ratings accounted for 22.2% of the mathematics performances (Desoete, 2007a). This suggests that teachers who are interested in metacognition in young children use multiple-method designs, including teacher questionnaires to get a complete picture of metacognitive skills.

Finally, a study reveals that the paradigm you choose is what you get (Desoete, 2007a). Child questionnaires do not seem to reflect actual skills, but they are useful to evaluate the metacognitive "knowledge" and "beliefs" of young children (Desoete, 2007a). Think-aloud protocol analyses were found to be accurate but time-consuming techniques to assess metacognitive "skills" of children with an adequate level of verbal fluency (e.g., Azevedo et al., 2007).

Training of Metacognition

We want to know in what ways we can enhance mathematical problem solving by focusing on metacognition. Hartman and Sternberg (1993) summarized the research literature in this field of the training of metacognition and presented four main approaches: promoting general awareness by modelling by teachers, improving metacognitive knowledge (knowledge of cognition), improving metacognitive skills (regulation of cognition), and fostering teaching environments. In the first approach of promoting general awareness, teachers *modelled* metacognitive skills and stimulated through a kind of reflective discourse the self-reflection exercises of students. Within the second approach, improving metacognitive knowledge, teachers handed out overviews of successful approaches/strategies that clarified how, when, and why to use specific strategies (e.g., Desoete & Roeyers, 2006). Children, for example, learned to slow down on more difficult tasks, to activate prior knowledge, to visualize and to build up diagrams. The third approach aimed to improve students' performance in mathematics by developing their metacognitive skills. This approach included presenting heuristics that are intended to support planning, monitoring, and evaluation. Domain-dependent skills or strategies focus explicitly on concepts that support the learning of a specific content. In contrast, domain-independent strategies are considered as general strategies. In the fourth type of approach, training researchers aim to create "powerful" teaching environments. These teaching environments foster self-reflection of children, improvement, and help students to attribute their success to the use of adequate strategies.

Mevarech and Kramarski's (1997) method, IMPROVE, is an example of metacognitive instruction that emphasizes reflective discourse by providing each student with the opportunity to be involved in mathematical reasoning. Teachers are trained to use metacognitive questions about the nature of the problem, the use of appropriate strategies for solving the problem and the relationships with previous knowledge (Kramarski, Mevarech, & Lieberman, 2001). IMPROVE is the acronym of all the teaching steps that constitute the method: Introducing the new material, Metacognitive questioning, Practicing, Reviewing, Obtaining mastery on higher and lower cognitive processes, Verification, and Enrichment and remedial. The metacognitive questioning includes four kinds of self-addressed questions: Comprehension question (e.g., What is the problem all about?), Connection question (e.g., How is the problem at hand similar to, or different from, problems you solved in the past? Please explain why), Strategic (e.g., What kinds of

strategies are appropriate for solving the problem, and why?), and finally, Reflection (e.g., Does the numeric solution make sense? Can the problem be solved in a different way?). This approach is in line with what Hartman (2001) identified as teaching “with” and “for” metacognition. This fourth type of method often combines informed strategy training (influencing metacognitive knowledge) and self-control training (influencing the metacognitive skills). Metacognitive-oriented training improved mathematical problem solving, even in adolescence (Montague, Woodward, & Bryant, 2004). To summarize, metacognitive skills were found to be trainable, even in adolescence. However, the metacognitive skills had to be explicitly taught to enhance mathematics.

From Cognitive and Metacognitive Deficits to STICORDI Devices

Children and adolescents with MLD need remediation techniques designed to address their cognitive (Proc. PT, Sem PT, Vis Sp PT, Num K FR in Figure 11.1) and eventual metacognitive (MP 4) deficits and make them less vulnerable.

It certainly is important to explicitly train metacognition if children with MLD have less developed metacognitive skills. This metacognitive training might improve the performance in mathematical problem solving. However, not all children with MLD have metacognitive deficits and metacognitive training therefore does not solve the problems of all children and adolescents with MLD.

STICORDI is an acronym referring to STI(mulation), CO(mpensation), R(emediation) and DI(spensation) (see Desoete, 2007b). STICORDI devices are needed if metacognition training programs fail. With STICORDI devices “reasonable” adjustments are provided to ensure that students with MLD are not placed at a substantial disadvantage compared to non-disabled peers. We give an example of how such a device, after an assessment of strengths and weaknesses, is applied to Silke.

Silke is a 13-year-old intelligent (TIQ 117, VIQ 122, PIQ 109) girl with MLD (MP 1 in Figure 11.1). The subtest scores on the WISC-III are Information SS = 13, Similarities SS = 15, Arithmetics SS = 7, Vocabulary SS = 15, Comprehension SS = 17, Digit Span SS = 12, Picture Completion SS = 12, Coding SS = 10, Picture Arrangement SS = 10, Block Design SS = 10, Object Assembly SS = 13, Symbol search SS = 10. At school, Silke’s performance is above average in reading and spelling, but poor in mathematics. The assessment reveals that Silke’s strengths are: size comparison, word-number transcoding and spatially representing numerical information in time-unlimited conditions (Vis Sp in Figure 11.1), insight into the structure of the small numbers (0 to 100), and a capacity for analyzing simple linguistic information (Num K in Figure 11.1). Silke’s weaknesses are fast retrieving mathematical facts (Sem PT in Figure 11.1) and sequencing multiple steps in complex procedures (Proc PT in Figure 11.1). She doubts her capacities and cannot predict the task difficulty very well (MP 4– in Figure 11.1). However she is very motivated (MP 2+ in Figure 11.1) and can work with concentration on mathematical problems (MP 3+ in Figure 11.1).

It is recommended that Silke receives comprehensive cognitive strategy instruction in written and mental calculation and training on metacognitive prediction skills. This intervention takes place, in close collaboration with the teacher, in a rehabilitation center twice a week in two 30-minute sessions for one year. In addition, the following STICORDI devices were used at school. Silke is provided with timely and structured handouts; she gets textbooks, structured charts and information sheets, so that she does not have to rely on what she copied during the lectures. She gets an individualized mathematics glossary of terms, concepts, information (number facts, working methods, mnemonics, etc.), and formulae. This booklet and a calculator can be used in examinations. She never has to

come unexpectedly in front of the class on the blackboard, answer a question or make a test. She gets 30% additional time for timed assessments and examinations in a special room, without disturbance from other students. STICORDI-devices as “reasonable” adjustments ensure that she is not placed at a substantial disadvantage compared to non-MLD disabled peers.

Conclusion

It is possible and necessary to obtain a picture of the efficient processes involved in mathematical problem solving of individuals with MLD, in order to analyze problem-solving mistakes. A profile should summarize strengths (MP1, 2, 3, 4) and weaknesses (Proc PT, Sem PT, Vis Sp PT, Num K PT) and facilitate instructional recommendations to be made. To sum up, individuals with MLD show shortcomings in different cognitive and metacognitive processes associated with mathematical problem solving. To focus on the particular problems of children and adolescents with MLD it is necessary to assess strengths and weaknesses. This assessment can easily be done in the classroom by a teacher in collaboration with a school psychologist. The student’s profile has several educational applications, aiding teachers and therapists in developing relevant instructional programs to optimize students’ cognitive recourses and creating environments that depend less on underlying automatic processes and poorly developed cognitive resources. STI(mulation), CO(mpensation), R(emediation) and DI(spensation) devices provide teachers with the necessary instructional expertise that will help them towards empowerment of vulnerable children and youth with MLD.

References

- Alexander, J. M., Carr, M. & Schwanenflugel, P. J. (1995). Development of metacognition in gifted children: Directions for future research. *Developmental Review*, 15, 1–37.
- Ansari, D., Donlan, C., Thomas, M. S. C., Ewin, S. A., Peen, T., & Karmiloff-Smith, A. (2003). What makes counting count? Verbal and visuo-spatial contributions to typical and atypical number development. *Journal Experimental Child Psychology*, 85, 50–62.
- Artzt, A. F. & Armour-Thomas, E. (1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. *Cognition and Instruction*, 9, 137–175.
- Azevedo, R., Greene, J. A., & Moos, D. C. (2007). The effect of a human agent’s external regulation upon college students’ hypermedia learning. *Metacognition and Learning*, 2(2/3), 67–87.
- Barbareis, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., & Jacobsen, S. J. (2005). Learning disorder: Incidence in a population-based birth cohort, 1976–82, Rochester, Minn. *Ambulatory Pediatrics*, 5, 281–289.
- Borkowski, J. G. & Thorpe, P. K. (1994). Self-Regulation and Motivation: A Life-Span Perspective on Underachievement. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 45–100). Hillsdale, NJ: Erlbaum.
- Brown, A. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms. In F. E. Weinert & R. Kluwe (Eds.), *Metacognition, motivation and understanding* (pp. 65–116). Hillsdale NJ: Erlbaum.
- Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (1998). Learning styles: A cross-sectional and longitudinal study in higher education. *British Journal of Educational Psychology*, 68, 427–441.
- Butterworth, B. (1999). *What counts: How every brain is hardwired for math*. New York: The Free Press.
- Cornoldi, C. & Lucangeli, D. (2004). Arithmetic education and learning disabilities in Italy. *Journal of Learning Disabilities*, 37, 42–49.
- Cornoldi, C., Venneri, A., Marconato, F., Molin, A., & Montinaro, C. (2003). A rapid screening

- measure for the identification of visuospatial learning disability in schools. *Journal of Learning Disabilities*, 36, 299–306.
- Cronbach, L. J. & Snow, R. E. (1977). *Aptitude and instructional methods: A handbook for research on interactions*. New York: Irvington.
- Davidson, G. R. & Freebody, P. R. (1986). Children and adults or novices and experts? A dilemma from cross-cultural developmental research. *Australian Journal of Psychology*, 38, 215–230.
- De Clercq, A., Desoete, A., & Roeyers (2000). EPA2000: A multilingual, programmable computer assessment of off-line metacognition in children with mathematical-learning disabilities. *Behavior Research Methods, Instruments & Computers*, 32, 304–311.
- Desoete, A. (2007a). Evaluating and improving the mathematics teaching-learning process through metacognition? *Electronic Journal of Research in Educational Psychology*, 5(3), 705–730.
- Desoete, A. (2007b). Students with mathematical disabilities in Belgium: from definition, classification and assessment to STICORDI devices. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in Learning and Behavioral Disabilities*, Vol. 20. *International Perspectives* (pp. 181–222). Amsterdam & Oxford: Elsevier Press.
- Desoete, A. (2008). Co-morbidity in mathematical learning disabilities: Rule or exception? *The Open rehabilitation Journal*, 1, 15–26.
- Desoete, A. & Grégoire, J. (2007). Numerical competence in young children and in children with mathematics learning disabilities *Learning and Individual Differences*, 16, 351–367.
- Desoete, A. & Roeyers, H. (2005). Cognition and metacognition in children with mathematics learning disabilities. In P. Ghesquière & A. J. J. M. Ruijsenaars (Eds.), *Learning Disabilities: A challenge to teaching and instruction*. *Studia Paedagogica* 40 (pp. 65–89). Leuven University Press.
- Desoete, A. & Roeyers, H. (2006). Current issues in the assessment and training of metacognition related to mathematical problem solving. In Alea V. Mitel (Ed.), *Focus on educational psychology* (pp. 251–275). Hauppauge NY: Nova Science Publishers.
- Desoete, A., Roeyers, H., & Buysse, A. (2001) Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities*, 34, 435–449.
- Desoete, A., Roeyers, H., Buysse, A., & De Clercq, A. (2002). Dynamic assessment of metacognitive skills in young children with mathematics learning disabilities. In D. Van der Aalsvoort, W. C. M. Resing, & A. J. J. M. Ruijsenaars (Eds.), *Learning Potential Assessment and Cognitive Training* (pp 307–333). JAI Press Inc/Elsevier Science Ltd.: England.
- Desoete, A., Roeyers, H., & De Clercq, A. (2004). Children with mathematics learning disabilities in Belgium. *Journal of Learning Disabilities*, 37, 50–61.
- Dowker, A. (2004). What works for children with mathematical difficulties. Research Report No 554. University of Oxford: UK.
- Dowker, A. (2005). *Individual differences in arithmetic. Implications for psychology, neuroscience and education*. Hove, UK: Psychology Press.
- Efklides, A. (2001). Metacognitive experiences in problem solving: Metacognition, motivation, and self-regulation. In A. Efklides, J. Kuhl, & R. M. Sorrentino (Eds.), *Trends and prospects in motivation research* (pp. 297–323). Dordrecht, The Netherlands: Kluwer.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of Intelligence*. Hillsdale, NJ: Erlbaum, pp. 231–235.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906–911.
- Focant, J., Grégoire, J., & Desoete, A. (2006). Goal-setting, planning and control strategies and arithmetical problem solving at grade 5. In A. Desoete & M. Veenman (Eds.), *Metacognition in mathematics education* (pp. 51–72). NOVA: New York.
- Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, 37, 4–15.
- Geary, D. C. & Brown, S. C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology*, 27, 398–406.
- Ginsburg, H. P. (1997). Mathematics learning disabilities: A view from developmental psychology. *Journal of Learning Disabilities*, 30, 20–33.

- Goos, M. (2002). Understanding metacognitive failure. *Journal of Mathematical Behavior*, 21, 283–302.
- Grégoire, J. (1997). Diagnostic assessment of learning disabilities. From assessment of performance to assessment of competence. *European Journal of Psychological Assessment*, 13, 10–20.
- Grégoire, J. (2005). Développement logique et compétences arithmétiques. Le modèle piagétien est-il toujours actuel? In M. Crahay, L. Verschaffel, E. De Corte, & J. Grégoire (Eds.), *Enseignement et apprentissage des mathématiques* (pp. 57–77). Brussels: De Boeck.
- Hartman, H. (2001). *Metacognition in learning and Instruction*. Dordrecht Netherlands: Kluwer.
- Hartman, H. & Sternberg, J. (1993). A broad BACEIS for improving thinking. *Instructional Science*, 21, 401–425.
- Kramarski, B. Mevarech, Z. R., & Lieberman, A. (2001). The Effects of Multilevel-versus Unilevel-Metacognitive Training on Mathematical Reasoning, *The Journal of Educational Research*, 94, 292–300.
- Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8–9-year-old students. *Cognition*, 93, 99–125.
- Lucangeli, D. & Cabrele, S. (2006). The relationship of metacognitive knowledge, skills and beliefs in children with and without mathematics learning disabilities. In A. Desoete & M. Veenman (Eds.), *Metacognition in mathematics education* (pp. 98–128). NOVA: New York.
- Lucangeli, D., Cornoldi, C., & Tellarini, M. (1998). Metacognition and learning disabilities in mathematics. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (pp. 219–285). Greenwich: JAI Press Inc.
- McNamara, B. E. (2007). *Learning disabilities: Bridging the gap between research and classroom practice*. New Jersey: Pearson Prentice Hall.
- Meltzer, L., Roditi, B., Houser, R. F., & Perlman, M. (1998). Perceptions of academic strategies and competence in students with learning disabilities. *Journal of learning disabilities*, 31, 437–451.
- Metcalfe, J. (Ed.). (1998). Personality and social psychology. Review (Special Issue). *Metacognition*, 2, 87–135.
- Mevarech, Z. R. & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal*, 34, 365–394.
- Miranda, A., Garcia, R., Marco, R., & Rosel, J. (2006). The role of the metacognitive beliefs system in learning disabilities in mathematics. Implications for intervention. In A. Desoete & M. Veenman (Eds.), *Metacognition in mathematics education* (pp. 152–171). NOVA: New York.
- Montague, M. (1998). Research on metacognition in special education. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (pp. 151–183). Oxford: Elsevier.
- Montague, M., Woodward, J., & Bryant, D. P. (2004). International perspectives on mathematics and learning disabilities: Introduction to the special issue. *Journal of Learning Disabilities*, 37, 2–3.
- Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Passolunghi, M. C. & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, 88, 348–367.
- Pressley, M. (2000). Development of grounded theories of complex cognitive processing: Exhaustive within- and between study analyses of thinking-aloud data. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 262–296). Lincoln, NE: Buros Institute of Mental Measurements.
- Pugalee, D. K. (2001). Writing, mathematics, and metacognition: looking for connections through students' work in mathematical problem solving. *School Science and Mathematics*, 101, 236–245.
- Reeye, R. A. & Reynolds, F. J. (2004, July). The nature of young children's mathematical (dis)abilities. Poster presented at the International Society of the Study of Behavioral Development. Gent.
- Schraw, G. (2000). Assessing metacognition: Implications of the Buros Symposium. In G. Schraw & J. C. Impara, (Eds), *Issues in the measurement of metacognition*, (pp. 297–321). University of Nebraska-Lincoln: Buros Institute of Mental instruments.

- Shalev, R., Manor, O., & Gros-Tsur, V. (2005). Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine and Child Neurology*, *47*, 121–125.
- Sperling, R. A., Howard, B. C., Staley, R., & Du Bois, N. (2004). Metacognition and regulation of cognition. *Contemporary Educational Psychology*, *27*, 51–79.
- Sternberg, R. J. (1985). *Beyond IQ. A triarchic theory of human intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R. J. (1990). *Metaphors of the mind: Conceptions of the nature of intelligence*. Cambridge: Cambridge University Press.
- Stock, P., Desoete, A., & Roeyers, H. (2006). Focussing on mathematical disabilities: A search for definition, classification and assessment. In Soren V. Randall (Ed.), *Learning disabilities new research* (pp. 29–62). Nova Science: Hauppauge, NY.
- Stock, P., Desoete, A., & Roeyers, H. (2007). Early markers for arithmetic difficulties. *Educational & Child Psychology*, *24*, 28–39.
- Tanner, H. & Jones, S. (2002). Assessing children's mathematical thinking in practical modelling situations. *Teaching Mathematics and its Applications*, *21*, 145–159.
- Thorpe, K. J. & Satterly, D. J. H. (1990). The development and inter-relationship of metacognitive components among primary school children. *Educational Psychology*, *10*, 5–21.
- Tobias, S. & Everson, H. (2000). Assessing metacognitive knowledge monitoring. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 147–222). Lincoln, NE: Buros Institute of Mental Measurements.
- Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 77–99). Münster: Waxmann.
- Veenman, M. V. J., Kok, R. & Blöte, A. W. (2005). The relation between intellectual and metacognitive skills at the onset of metacognitive skill development. *Instructional Science*, *33*, 193–211.
- Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction*, *14*, 89–109.
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning. Conceptual and methodological considerations. *Metacognition Learning*, *1*, 3–14.
- Verschaffel, L. (1999). Realistic mathematical modelling and problem solving in the upper elementary school: Analysis and improvement. In J. H. M. Hamers, J. E. H. Van Luit & B. Csapo (Eds.), *Teaching and learning thinking skills. Contexts of learning* (pp. 215–240). Lisse: Swets & Zeitlinger.
- von Aster, M. (2000). Developmental cognitive neuropsychology of number processing and calculation: Varieties of developmental dyscalculia. *European Child and Adolescent Psychiatry*, *9*, II/41–II/57.
- Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology*, *89*, 1–14.
- Winne, P. H. & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. E. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego: Academic Press.
- Wong, B. Y. L. (1996). Metacognition and learning disabilities. In B. Y. L. Wong (Ed.), *The ABCs of learning disabilities* (pp. 120–140). San Diego: Academic Press.
- Zimmerman, B. J. (2000). Attaining Self-Regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation: Theory, research, and applications* (pp. 13–41). San Diego, CA: Academic Press.

Part VI

Individual Differences

12 Context Matters

Gender and Cross-Cultural Differences in Confidence

Mary Lundeberg and Lindsey Mohan
Michigan State University

Many centuries ago, the Chinese spiritual leader and philosopher Confucius stated the potential educational import of what we term confidence discrimination, and it may be of even greater import today: “When you know a thing, to hold that you know it; and when you do not know a thing, to allow that you do not know it: That is knowledge”.

(Burton, 1967, p. 1060)

Confidence in one’s knowledge is a form of metacognition important in academic settings. Learners must be able to accurately judge their knowledge in order to know where to focus their efforts when they study. If students are confident in their understanding, they can move on to more difficult concepts; so assessing whether one knows something or not is an important part of school life.

However, what students think they understand may not be understood. This has instructional implications: Teachers, too, need to be aware of students’ confidence in their understanding and provide opportunities for students to compare their confidence with their actual understanding. If students express high levels of confidence in misconceptions, teachers may need to spend additional instructional time for students to relearn these concepts. Inappropriate judgments of confidence may interfere with student effort; students may not pay attention if they believe they already know something. Moreover, if wrong answers are agreed upon in small group settings, these concepts prove to be even more difficult to re-teach, because two heads are usually more confident than one, even when the answer is wrong (Puncochar & Fox, 2004).

How are gender and cross-cultural differences in confidence related to metacognition? Lack of confidence has been cited frequently as the reason that girls do not persist in science and mathematics, and that women do not persist in higher education. Although most young children are quite confident in their abilities, even first- and second-grade boys express higher expectations about their academic abilities than do girls (Pressley & Ghatala, 1989). Based on NAEP data, girls at fourth, eighth and twelfth grade report lower confidence than boys when asked to respond to the statement “I am good at math” (McGraw, Lubienski, & Strutchens, 2006). Gender differences in self-confidence increase with age (Hyde, Fennema, Ryan, Frost, & Hopp, 1990); even when female students achieve at equal or higher levels than their male peers, they underestimate their abilities (Zukerman, 1987). Similarly, some cultures may have a tendency towards overconfidence and other cultures towards humility. The Programme for International Student Assessment (PISA) measured both mathematical performance and beliefs related to mathematical performance, such as students’ confidence in their ability to do mathematics (self-concept) and students’ capacity to overcome learning challenges that they find difficult (self-efficacy). While the United States is rated 25th out of 40 countries on this

recent international mathematics achievement assessment (PISA, 2003), Korea is ranked first. Even with the dismal mathematics performance scores, the United States has the distinction of having students with the strongest self-concept and self-efficacy scores, as compared to Korea, where students had the weakest self-concept and self-efficacy scores. So although American students do not score as well as the Koreans, they have more confidence in their mathematics abilities. This pattern of cultural difference in confidence and achievement in mathematics between Korean and American students was also found in international assessments given in 1990 (Krauthammer, 1990). In this assessment, Koreans also ranked first in mathematics achievement and students in the United States ranked last on achievement; ironically, the reverse was true with the confidence measure (e.g., “I am good at mathematics”). However, most gender and cultural studies have measured general confidence, rather than confidence that is specific to a given task or item.

In this chapter, we first analyze the ways researchers have measured confidence and variables such as gender and culture. After summarizing results of studies of gender and culture on confidence, we argue that context matters in studies of gender and cultural differences in confidence, and girls and boys become influenced by norms of behavior in their classrooms, cultures, and in their countries. We provide examples to illustrate the role of gender, culture, and confidence, and finally, conclude by providing suggestions for educators concerned about cultivating appropriate confidence in their classroom contexts.

Measuring Confidence

As others in this handbook have discussed (e.g., Schraw; Tobias & Evenson), researchers have many ways of measuring confidence, that is, whether someone thinks they know something or not. We distinguish item-specific confidence from global confidence, discuss ways to measure item-specific confidence, and ways to assess confidence discrimination. Finally, we discuss our rationale for measuring confidence in classroom settings, and within those settings, how researchers have analyzed culture and gender differences.

Kind of Confidence: Global or Item Specific

Most researchers studying cultural or gender differences have examined global feelings of confidence. Global confidence is similar to self-concept; this type of confidence asks about a global feeling of “being good at school” or global feeling about a specific domain such as “being good at writing.” In such studies (e.g., Casey, Nuttall, & Pezaris, 2001) students are asked to make global judgments (e.g., “How good are you at mathematics?”). Gender differences in confidence, especially in mathematics, are fairly consistent and generally show that men are more confident than women when asked to report general feelings of confidence.

In contrast, performance can also be tied to a particular test item or specific task: This type of confidence requires students to judge whether they could or did complete a specific item or task accurately. This type of confidence is very specific to a given item or task and is typically asked prior to trying a task (i.e., hypothetically) or after the student has completed the task. For example, some researchers have asked students hypothetical questions about self-efficacy that are tied closely to a particular specific context, such as “How confident are you that you can solve X math problem?” (Pajares, 1996), or “How confident are you that you can earn an A, B, C, D in science class?” (Britner & Pajares, 2001). When researchers focused on specific items, students’ hypothetical judgments may be later correlated with actual performance or grades; for example, Lloyd, Walsh, and Yilagh (2005) asked students to rate hypothetical confidence on answering mathematical problems and

students then completed the problems two weeks later as a measure of mathematical achievement. Girls' achievement on the math problems met or exceeded boys' achievement; however, girls were more likely to show underconfidence and boys were more likely to be overconfident. This corroborates findings by Jones and Jones (1989) who compared hypothetical judgments to subsequent performance and found that contextual differences affect gender. They found gender differences when considering the ability level of the student and the types of questions asked, reporting that females had lower confidence judgments compared to males, particularly on two science items and an unfamiliar math question. Interestingly, they found high-ability females to report the lowest overall confidence judgments for science and unfamiliar math items, but the highest overall confidence on the familiar math item.

Much less is known about situations in which students first do a problem and then report how confident they are that they solved the problem (or answered the question) accurately. This kind of item-specific confidence is sometimes called meta-memory, or feeling of knowing. Having an objective measure, such as performance, along with a subjective measure of confidence, allows researchers to calibrate the relationship between the accuracy of a set of responses and students' metacognitive assessment (i.e., their confidence that their response was correct). When students express higher confidence than warranted given the accuracy of responses, they are said to have a positive response bias (e.g., Stankov & Crawford, 1996). Researchers have found that people are generally overconfident about their cognitive abilities—that is, people believe their responses to general knowledge items, vocabulary items, and standardized test questions are more accurate than the responses actually are (Liberman & Tversky, 1993; Kleitman & Stankov, 2001).

Furthermore, students estimate their confidence more accurately *after* they have answered an item, rather than *before* (as in a hypothetical situation) (e.g., Glenberg & Epstein, 1987). Because students tend to be overconfident about their understanding, particularly before they test their understanding, researchers studying item-specific confidence should first ask students to perform a given task (such as answer an item) and then assess confidence. Similarly, in a classroom situation, teachers should not ask at the end of a lecture, "Does anyone have any questions?" Rather, if instructors truly want to assess students' confidence in understanding, they should give students a task or problem to perform and then assess confidence.

Ways to Examine Item-Specific Confidence: Calibration or Discrimination

In Table 12.1 we summarize studies that have used item-specific confidence judgments. Measuring item-specific confidence allows researchers to *calibrate* confidence—to compare students' confidence ratings with how often they are correct (frequency). As we explained earlier:

How well-calibrated an individual's confidence estimates are can be assessed by the relationship between accuracy and confidence judgments. A person is perfectly calibrated when, across all judgments at a given level of confidence, the proportion of confidence judgments is identical to the actual probability of being correct (Lichtenstein, Fischhoff, & Phillips, 1977) . . . Confidence ratings typically fall below the ideal calibration function, indicating overconfidence, and usually depart from ideal calibration accuracy by 19% to 20%. (Lundeberg, Fox, Brown, & Elbedour, 2000, p. 153)

Table 12.1 Studies of Confidence, Gender and Culture

<i>Author</i>	<i>Age/location</i>	<i>Content</i>	<i>Type of confidence</i>	<i>Type of items</i>	<i>Context</i>
Koivula, Hassmen, & Hunt (2001)	High-school seniors; college freshmen (18–19 year olds), USA	SweSAT performance: verbal and quantitative sections	Item-specific	Self-assessments of confidence made immediately after answering questions on verbal and quantitative reasoning subtests of the SweSAT Bem Sex Role Inventory (BSRI)	Experimental
Lundeberg, Fox, & Puncocochar (1994)	Undergraduate and graduate, USA	Laboratory methods in psychology; memory course	Item-specific	Multiple-choice items on final exam measured computation, experimental design, statistics and conceptual content; MC and TF in human learning and memory. After answering item, students rated confidence on a 5-point scale, with 1 = pure guess and 5 = very certain	Classroom
Jonsson & Allwood (2001)	Grades 11 and 12, Sweden	SweSAT: Word knowledge (WORD/verbal) and logical/spatial ability (DTK)	Item-specific	<i>SweSAT verbal (WORD) and spatial/logical (DTK)</i> : Rate confidence on each item after completing it. Administered three times <i>Need-for-cognition test</i> : used to determine difference between “academics” and “workers”	Experimental
Pallier (2003)	STUDY 1: college freshmen, Australia	Psychology course	Item-specific	General knowledge test: 20 questions about history, science, geography, technology, and current events on computer Visualization test: line length test on computer Answered the questions then rated their confidence on a 20–100% scale (20% means guessing and 100% was absolutely sure)	Experimental

STUDY 2: general population, Australia	None	Item-specific	General knowledge test, Synonyms Vocabulary test, Letter Series Test, Cattell's Matrices Test	Experimental
Lloyd, Walsh, & Yilagh (2005)	Math	Hypothetical	<i>Math self-efficacy</i> : nine math problems to rate hypothetical confidence (students completed problems 2 weeks later)	Experimental, classroom
Jones & Jones (1989)	Science and math (familiar and unfamiliar items)	Hypothetical	Four items Told subjects to decide whether you will get answer right and did problems later	Experimental
Yates, Lee, & Shinotsuka (1996)	General	Item-specific	Two general-knowledge items that they rated confidence on THEN explicitly explained confidence judgments to participants and participants responded whether they thought Taiwan, American or both would show overconfidence.	Experimental
Lundeberg, Fox, Brown, & Elbedour (2000)	Mathematics, social science, science, and speech courses	Item-specific	Course exams; immediately after answering each item students rated confidence that answer on each item was correct on scale of 0% (uncertain) to 100% (correct)	Classroom

A second way to examine student confidence is to estimate whether students express high feelings of confidence when appropriate and lower feelings of confidence when uncertain, that is to show *confidence discrimination*. This way of examining student confidence is even more important in an educational setting, to show more confidence when correct and adjust confidence when incorrect. As we explained:

Confidence discrimination is determined by examining whether people are able to discriminate between what they know and what they do not know. We would expect students to assign higher confidence for correct answers than for incorrect answers. The conceptual distinction between confidence calibration and discrimination is an important one: “Good calibration does not necessarily entail good discrimination” (Yanniv, Yates, & Smith, 1991, p. 613). Clearly, being overconfident when wrong may not be a desirable trait in many situations. Good confidence discrimination, especially knowing when one does not know something, is a critical topic in an educational setting. (Lundeberg et al., 2000, p. 153)

Ways to Assess Discrimination: Significance Tests or CAQs

We have calculated confidence discrimination in earlier work by examining the differences when students answered correctly compared to when they answered incorrectly, and developed what we referred to as a confidence accuracy quotient (CAQ). Essentially the CAQ calculations allow researchers to “standardize individuals’ mean confidence ratings by accuracy according to the following formula: $CAQ = (\text{mean confidence correct} - \text{mean confidence wrong}) / \text{the overall standard deviation of individual confidence scores}$ (Shaughnessy, 1979)” (Lundeberg et al., 2000, p. 155). A positive CAQ indicates that an individual reports higher confidence when they are correct, while a CAQ score of zero indicates the participant does not discriminate between instances when they are correct compared to incorrect.

A third way to assess item-specific confidence is to ask groups to rate their own confidence and then estimate how likely they think groups from another context would be overconfident. For example, Yates, Lee, and Shinotsuka (1996) asked students to rate their confidence to two general knowledge items and then had them respond whether they thought Taiwanese or American or both would show overconfidence. Taiwanese participants were twice as likely as Americans to think Americans would show overconfidence.

Why are people overconfident in their knowledge? Some researchers think that humans have general cognitive biases and heuristics that mediate judgments and tend to lead to errors (e.g., Kahneman & Tversky, 1996). Others believe that because general-knowledge type items are sometimes misleading, people are tricked into thinking they are answering correctly. In contrast to bias theory, which posits that overconfidence is derived from individuals, this Ecological Approach suggests that overconfidence comes from test items—that it is related to cues in the environment that result in miscalibration of confidence (Pallier, 2003). However, asking students about general knowledge may be less ecologically valid than an actual classroom situation, because they are familiar with the classroom situation.

Contexts for Measuring Confidence: Experimental or Classroom Setting

As shown in Table 12.1, only a few researchers examine item-specific confidence in actual classroom contexts by measuring classroom learning with teacher-designed tests

(Lundeberg et al., 2000; Lundeberg, Fox, & Puncocchar, 1994). Most research has been conducted in experimental contexts, measuring standardized exam performance or measuring general knowledge. Our rationale for conducting research on confidence in classrooms is that “. . . at some point, research bearing on instructional and learning issues in the classroom needs to be carried out systematically in the classroom” (Lundeberg et al., 2000, p. 158). Classroom research is more complicated in that it is difficult to control for variables; however, given that classroom and lab studies of similar questions have produced different results (Lundeberg & Fox, 1991), it is important to analyze questions in actual classroom settings. In our previous work, we looked at a variety of educational settings, including multiple courses taught by different course instructors in different countries. We tried to reduce the potential of confounding variables by grouping courses based on the overlap of the course content, so that courses taught in different countries might be grouped given a high degree of similarity in the content covered. We were concerned about controlling for the diversity of course instructors, but concluded that in cross-cultural research, it was more important that courses be taught by someone familiar with the culture, rather than taught by an outsider. Thus, in classroom-based research, such as our previous work, attending to ecological validity was a critical concern for researchers.

Measuring Culture: By Country, by Ethnicity

Some researchers (Heine, Kitayama, & Lehman, 2001; Lundeberg et al., 2000; Yates et al., 1996), have used “country” as a proxy for “culture”—that is, we assumed that people from different countries have different cultures. Of course, there are also ethnicities and cultural variations within each country, so “country” is a rough measure of culture. Within the same country, researchers have also examined the degree to which race (e.g., Hispanic, White) affects cognition and motivation (e.g., Stevens, Olivárez, & Hamman, 2006). Other researchers (e.g., Klassen, 2004) have used father’s country of birth as an estimate of culture, categorizing students into “Anglo Canadian” if students had fathers from Canadian, American, or European backgrounds and “Indo Canadian” if students’ fathers were Punjabi Sikh immigrants. Finally, some researchers have used scales to measure culture, for example Kim and Omizo (2005) used an Asian Values Scale and a European American Values Scale for Asian Americans to group students according to cultural values, recognizing that acculturation and enculturation affect the relationship between racial and ethnic identity.

Measuring Gender: By Sex, by Instrument

Similarly, researchers studying gender differences have used “sex” as a proxy for “gender,” assuming that participation in a group leads to social and cultural norms that influence women and men. Thus, the American Psychological Association (APA) recommends that “sex” may be misinterpreted as sexual activity and encourages researchers to use the term “gender” (APA, 2007). Most researchers agree that masculinity and femininity vary depending on the context and time, are not unidimensional constructs but orthogonal variables representing two distinct dimensions (Pajares & Valiante, 2001). For example, Pajares and Valiante included a measure of gender orientation, rather than just sex, on the basis that gender differences may be related to “the stereotypic beliefs about gender that students hold” (p. 368). The measure they used was based on items used in Harter and colleagues’ studies (Harter et al., 1997) and was adapted from the Children’s Sex Role Inventory (CSRI; Boldizar, 1991). Sample items assessing masculinity include questions such as “I like building and fixing things”; those assessing femininity include

items such as “I am a warm person and express those feelings to those I feel close to” (p. 371).

Gender and Cross-Cultural Differences in Confidence

In the studies using item-specific confidence summarized in Table 12.1, men showed signs of over-confidence. However, context matters greatly in whether gender differences emerge. These contexts include how confidence is measured (overall confidence as compared to confidence discrimination), the subject area (mathematics vs. other areas), characteristics of the subjects (levels of students, stereotypical gender traits), and culture.

Confidence Discrimination Matters

When overall confidence is examined, irrespective of correct and incorrect responses, men may appear to be more confident than women, or both women and men may show signs of being more confident than warranted by their accuracy. For example, Lloyd et al. (2005) examined the hypothetical confidence and mathematics achievement on a standardized test of fourth-grade and seventh-grade girls and boys in Canada. They calculated an overall self-efficacy score and compared that to the achievement scores for each sex, and found that although girls’ achievement was equal to boys, girls were “underconfident.” They concluded that “. . . educators, parents, and researchers alike should be concerned about the finding that girls tend to be underconfident of their mathematics ability relative to their actual achievement, whereas boys tend toward overconfidence: Even though girls’ mathematics achievement was commensurate with that of boys, this study suggests that girls do not yet know how able they are in mathematics. In a sense, they are good at mathematics, but do not know it” (p. 405). Unfortunately, these researchers did not examine confidence when girls and boys were incorrect as compared to when they were correct—if they had examined confidence discrimination, they may have reached a slightly different conclusion.

Gender differences occur most clearly when confidence discrimination (sometimes called resolution) is examined, that is, do students know what they know and what they do not know? When females are asked to report item-specific confidence judgments on correct items, they are equally or more confident than males, contrary to the stereotype of women having less confidence than men (Koivula, Hassmen, & Hunt, 2001; Lundeberg et al., 1994, 2000; Pressley, Levin, Ghatala, & Ahmad, 1987). When incorrect, women report lower confidence, whereas men tend to be overconfident (Koivula et al., 2001; Lundeberg et al., 1994; Pressley et al., 1987). Thus, females seem to be more accurate than males in calibrating their confidence—their metacognition is more aligned with their actual performance. This trend is shown in Figure 12.1. In explaining the results of Figure 12.1, Lundeberg et al. (1994) wrote:

In general, when men were wrong, their confidence was close to 4 on the 5-point scale (reasonably certain), whereas when women were wrong, their confidence was closer to 3 on the scale (mixed feelings of confidence and uncertainty). Overall then, although both groups were overconfident, men were consistently more confident than they should have been when they were wrong. Moreover, women showed much greater tendency to calibrate their confidence than did men.” (p. 118)

Similarly, on the mathematics subtest of the Swedish version of the SAT, Koivula and colleagues (2001) found that men rated themselves more “extremely sure” when correct

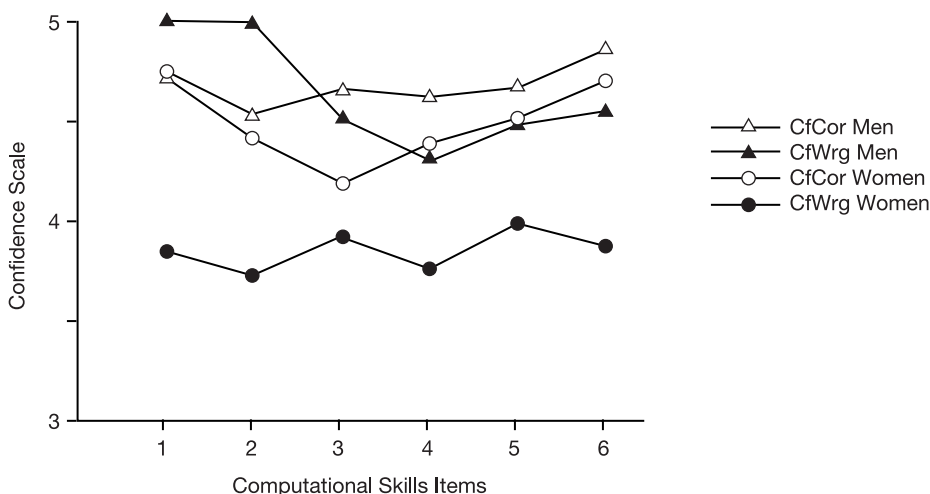


Figure 12.1 Lab 1 mean confidence of women and men on computational skill items.

Note: CfCor Men = confidence when correct for men; CfWrg Men = confidence when wrong for men; CfCor Women = confidence when correct for women; CfWrg Women = confidence when wrong for women.

than did women, and more importantly, men rated themselves more “extremely sure” when they were incorrect than did women.

Subject Area Matters

This trend in gender differences and confidence, however, was not observed for all disciplinary content (Jonsson & Allwood, 2001; Koivula et al., 2001; Lundeberg et al., 1994). Although gender differences occasionally appear on writing tasks (Pajares & Valiante, 2001), such effects show up more often with mathematics than with other subject matter.

For example, gender differences in confidence were most pronounced on the computational items, and were not present on the experimental design items in our research (Lundeberg et al., 1994). Koivula and colleagues (2001) examined both performance of Swedish high school and college students on the SweSAT (the Swedish version of the SAT) for both mathematics and a word subtests and found no differences in the verbal subtest for confidence ratings or for achievement, but did find gender differences in confidence for mathematics (as reported in the previous section). Jones and Jones (1989) measured confidence in high-school students on four items (two science and two mathematical; unfamiliar and familiar). Females were more confident than males on the familiar (computational) mathematics items, but less confident than males on science questions and the unfamiliar mathematics items.

Patterns of overconfidence in men that were dependent on the context of the subject were also evident in Australian college students using tests of cognitive ability in two studies (Pallier, 2003). In general, men were more confident than women on general knowledge items, although both were overconfident. Women were more underconfident than men on the visualization test, although both sexes were underconfident. In his second study, Pallier (2003) found that men and women were overconfident on all four tests, although men were significantly more overconfident than women on vocabulary, letter series, and cattell matrices. Unfortunately, Pallier did not calculate confidence

discrimination (resolution) scores, so we cannot interpret these findings in light of whether the subjects' confidence varied when they were correct or incorrect.

Jonsson and Allwood (2001) examined the confidence of Swedish high-school seniors using older versions (1993) of the SweSAT tests of word knowledge and logical/spatial ability, and administered one-third of these tests three times (within a six-week period), so they could examine whether confidence scores were stable across individuals at different points in time. These researchers found no gender differences in confidence in logical/spatial ability, but they did find gender differences in the brief 15-minute test of word knowledge. In general, males were more accurate/realistic in confidence at Times 1 and 2, but overconfident at Time 3, and females were underconfident at Times 1 and 2 but more realistic at Time 3.

Beyer (1999) did not calculate item-specific confidence, but did examine gender differences in the accuracy of grade predictions for college students in Introductory Psychology, Social Psychology, and Computer Science. She found that women overestimated their grades less than men in Introductory Psychology (they were more accurate at predicting scores on upcoming tests than were men); however, there were no gender differences in the other two courses. Although in our own work, we found subject area differences in math items in 1994, we did not find gender differences in confidence across subject areas in our study six years later (Lundeberg et al., 2000), even though we had mathematics courses (e.g., Business Calculus, Calculus 2) and subjects involving language (e.g., Speech, Content Area Reading, Speech Theory). We concur with Jonsson and Allwood (2001) that gender differences on confidence appear to be unstable over time and dependent on knowledge domain and specific tasks.

Characteristics of Populations Matter

Trends regarding male overconfidence have been most pronounced in undergraduate populations (as compared to graduate students), introductory courses (as compared with upper-level courses), and with students in the lower quartile of class (as compared to the group in the top quartile) (Beyer, 1999; Jones & Jones, 1989; Lundeberg et al., 1994). In classroom studies, male students with lower ability (Jones & Jones, 1989; Lundeberg et al., 1994) and less experience in school (Beyer, 1999; Lundeberg et al., 1994) show more overconfidence than females. Pajares (1996) found that while eighth-grade students were generally biased toward overconfidence on math performance, gifted students were more accurate in their self-perceptions of confidence, and gifted girls were biased toward underconfidence. However, gender differences in confidence are not apparent in graduate students and juniors and seniors (Beyer, 1999; Lundeberg et al., 1994).

Although most researchers have equated sex with gender, some researchers have measured gender using a sex role inventory (Koivula et al., 2001; Pajares & Valliante, 2001). Koivula et al. (2001) measured gender using the Bem Sex Role Inventory (BSRI), a self-report instrument that categorizes individuals based on instrumental and expressive traits that are stereotypically considered male or female. The BSRI groups people into four categories: gender-typed (women or men who display responses considered gender-appropriate), cross-gender typed (women who display male characteristics and vice versa), androgynous (having both male and female traits), and undifferentiated. Categories on the BSRI mattered only in the correct context—that is, gender-typed men and cross-gender typed women displayed higher confidence than undifferentiated groups, and androgynous women and men had higher confidence than gender-typed women, cross-gender typed men, and undifferentiated women and men. When men were incorrect, they were highly confident regardless of categories on the BSRI, and when women were incorrect they

displayed lower confidence, again, irrespective of their score on the BSRI. These authors argue for including some measures of gender orientation in future studies; however, there are construct validity issues to consider when using such inventories: What does it really mean to be androgynous?

Other researchers have found that a feminine orientation (not necessarily sex) is adaptive for writing self-efficacy, writing self-concept, value for writing, task goal orientation, writing achievement, etc. (Pajares & Valiante, 2001). Thus, these researchers may have reached different conclusions had they been grouping students by sex only. Therefore, gender typing is another characteristic of students that matters in some situations.

Culture Matters

Fewer researchers have focused on cultural aspects of confidence, especially those who examine both gender and cultural differences. As shown in Figure 12.2, Lundeberg et al. (2000) found clear patterns of confidence discrimination among three cultures (Asian, Palestinian and US). As we explained earlier:

As indicated in the top panel of Figure 2 (for mathematics), students in Taiwan showed the greatest discrimination in confidence between correct and incorrect responses, whereas the Palestinian students showed the least resolution. This pattern of confidence discrimination (or lack thereof, such as in the Palestinian data) is likewise evident in the domain of social science courses . . . In summary, the most striking cultural difference obtained was in confidence discrimination, with the Palestinian students showing little discrimination in confidence when they were correct versus incorrect on class examination items. (Lundeberg et al., 2000, p. 156)

This pattern of better confidence discrimination for Asian cultures and poorer confidence discrimination for Middle Eastern cultures has been corroborated in a few other studies as well. For example, Puncochar, Fox, Fritz, and Elbedour (1996) found Israeli students to be more overconfident than students in the United States.

Most cross-cultural studies of confidence use general-knowledge questions rather than questions from actual classroom tests. Using general-knowledge questions, Yates et al. (1989) found that Chinese participants showed better confidence discrimination than did people in the United States and Japan. In a later study, Yates et al. (1996) asked students to rate their confidence to two general-knowledge items that they answered and then explicitly explained confidence judgments to participants and had participants respond whether they thought Taiwan, American, or both would show overconfidence (or high tendency judgments). Over 62% of the participants thought Americans would show more overconfidence compared to Taiwanese. Of the participants that said Taiwanese would show higher overconfidence, the Taiwanese participants were more than twice as likely to say this compared to Americans. Thirty percent of Americans said that the two countries would be equally confident. Just under half of the participants, both Taiwanese and American, explained overconfidence as a “disposition” that is influenced by the country in which one lives.

Culture is also associated with race. Stevens et al. (2006) also used global measures of confidence to examine racial differences between Hispanic and other students in Texas. They found Hispanic students had lower math achievement and reported lower feelings of self-efficacy and higher math anxiety compared to white students. However, they reported higher math interest and intrinsic motivation and reported knowing more people who were models for mathematics performance.

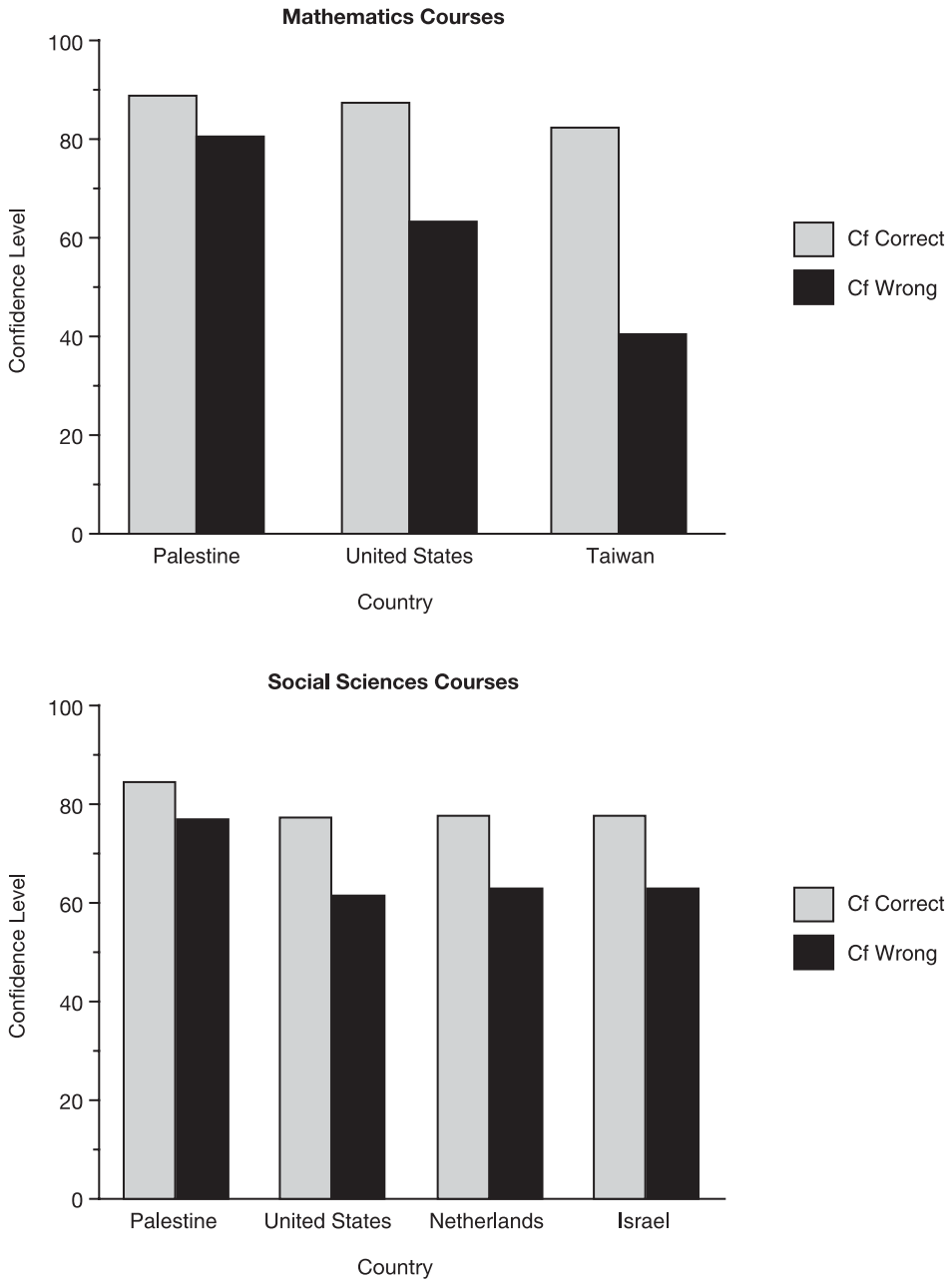


Figure 12.2 Confidence discrimination in mathematics courses in Taiwan, the United States, and Palestine, and in social science courses in Israel, the United States, Palestine, and the Netherlands. Cf = confidence.

Context of the Research is Likely to Matter

Most of the researchers studying item-specific cross-cultural confidence have done so in the context of a laboratory study—that is, using test items that are either modified from standardized test questions or general-knowledge questions—rather than using tests that

students are taking within a classroom context. Interestingly, different results emerge with regard to Asian populations in the classroom and lab research contexts. For example, lab researchers using general knowledge (almanac-type questions, such as “Which is longer (a) Panama Canal or (b) Suez Canal?”) have found that Asian participants are more overconfident than their Western counterparts (Wright et al., 1978; Wright & Phillips, 1980). In these studies, general-knowledge questions have been *translated* into Chinese, and students in Western countries have generally performed better on these items, which raises the question of cultural bias of the items.

Using actual course questions (e.g., Lundeberg et al., 2000) rather than general-knowledge test items may eliminate some of the bias that may have occurred in these former laboratory studies by Wright and colleagues. How much does the “authentic” context of the classroom matter? A meta-analysis comparing classroom and lab studies on essentially the same research question (test expectancies) found different empirical results (Lundeberg & Fox, 1991). Although not enough classroom research has been published to do a meta-analysis on culture and item-specific confidence, the differences found regarding Chinese calibration and confidence resolution may be related to the effects of the classroom as compared to lab research contexts.

Cultural Influences on Gender and Confidence

Gender differences in item-specific confidence tended to be quite small when viewed within countries (e.g., Lundeberg et al., 2000). Six years prior to studying cultural differences, we found women were as confident as men when they were correct (Lundeberg et al., 1994). Yet these differences were minimal when viewed within countries, and there was much greater variability across cultures than within cultures.

Other studies examining culture and gender have not used item-specific confidence measures. Two studies, however, show no gender or culture related differences (Klassen, 2004; Stetsenko, Little, Gordeeva, Grasshof, & Oettingen, 2000) whereas a third indicates that context matters in gender differences. For example, Saunders, Davis, Williams, and Williams (2004) found that African American females reported significantly higher academic self-efficacy, stronger intentions to complete the school year and had significantly higher GPAs than did African American males.

In a cross-national study of high-school students looking at mathematics performance and global confidence (e.g., “I am good at mathematics”), there were several differences across gender and country (Ercikan, McCreith, & Lapointe, 2005). Confidence was the strongest predictor of math achievement for both males and females in Norway and Canada and males in the United States, but NOT females in the United States. The strongest predictor for US girls’ math achievement was parent’s highest level of education, but confidence was the second highest predictor for US girls. In Canada, confidence was a stronger predictor for males than females, while the opposite was true in Norway.

Overall the evidence indicates that context matters with regard to gender and cultural differences in confidence. Moreover, this is important because confidence may be related to achievement, as well as metacognition (Sabina & Lazar, 2005). How one measures the confidence (item-specific) as compared to global sometimes produces different results.

Examples and Implications for Teachers

So what are the implications of this research for teachers and practitioners? In Table 12.2, we list suggestions for teachers, based on the research on confidence, gender, and culture.

Table 12.2 Ten Principles for Classroom Practice based on Implications of this Research*Classroom level*

1. Give students opportunities to assess their confidence on classroom tests.
2. Give students opportunities to assess their confidence on performance tasks, such as writing, small group cooperative projects, or paired performances (e.g., authoring web posters).
3. Assist students in creating rating scales for open-ended performance tasks.
4. Discuss students' and class's confidence as related to performance and encourage students to set and monitor goals for confidence calibration and discrimination (resolution).
5. Re-teach concepts if students are highly confident, but mistaken.
6. Discuss gender and cultural stereotypes with students, especially as they appear in curriculum materials, media, group interactions (boys dominate talk), and classroom tasks (e.g., girl as notetaker, boy as doer).
7. Give students opportunities to question stereotypes (Francis & Skelton, 2005).
8. Monitor and intervene in situations involving stereotypical behaviors.

School, district level, or in education courses with teachers

9. Discuss at the institutional level: *How is difference recognized and valued within the school?* (Francis & Skelton, 2005, p. 150).
10. Give teachers questions to discuss regarding attitudes and behavior (Francis & Skelton, 2005, pp. 139–140):
 - *Do I expect pupils to act differently because they are boys or girls?*
 - *Do I have different expectations of their abilities or potential based on whether they are a boy or a girl?*
 - *To what extent do I differentiate between the children? Am I aware that “girls” and “boys” are not homogeneous groups; for example, some girls will act in assertive ways and some boys will work cooperatively?*
 - *How do my interactions influence pupils and their understanding of what it means to be a girl or a boy?*
 - *What aspects of my daily work practices help pupils to understand themselves and others as girls and boys?*
 - *What are the main theories and ideas about teaching and learning that define my work with pupils? Do these take account of gender?*

Most suggestions focus on classroom teachers; the last two suggestions are for school administrators or for faculty teaching inservice or preservice teachers.

Some researchers have found that performance improves if students are given the opportunity to assess their confidence on tests. For example, Koivula et al. (2001) found that self-assessment of confidence improved the performance of both women and men as compared to students who took the same test but did not engage in thinking about how confident they were that their answer to each item was correct. This self-assessment of confidence was more beneficial especially for gender-typed women, cross gender-typed men, and androgynous men and women, as compared to the groups who did not do self-assessment—that is, thinking about whether one's answer was accurate or not led some students to get more correct answers. Gender-typed men and cross-gender typed women were less metacognitively aware of their performance, so self-assessment did not help their performance. However, Beyer (1999) found that students can improve in metacognitive awareness of the accuracy of their responses over time. This seems especially important for male students, who tend toward overconfidence in inaccurate answers (at least in some studies). Thus, we recommend that teachers ask students to estimate confidence and check confidence calibration.

Teachers can ask students to estimate confidence both in test situations and also on performance tasks. For example, after answering each question on a test, students can indicate the level of confidence they feel that their answer is correct or incorrect on a scale

of 1 (pure guess) to 5 (very certain). On individual exams, students can then later examine the incorrect items that they felt confident were correct and were mistaken, as well as correct items that they were unsure of. If students first take an exam individually, and then compare responses with a small group of team members (as in a cooperative exam situation), the teacher can ask small groups of students to come to consensus on the correct response, taking into account students' confidence.

For more open-ended tasks, such as group projects, multimedia performances or writing tasks, teachers can involve students in creating rubrics so they clearly understand the criteria on which they will be evaluated. Prior to turning in such an assignment, students can then use the rubric to evaluate their own work and their peer's work. Students can also indicate the confidence they feel in evaluating the work. This indicates how well students have internalized the criteria for a given performance. Such rating tasks assist especially lower-performing students to better understand expectations and can improve students' writing performance (Troia, 2007). Collaborative writing in which students evaluate one another's work has a strong impact (effect size of 0.75) on improving student writing (Graham & Perin, 2007).

Discussions with students regarding their individual confidence calibration and resolution would seem to be helpful, and there is some indication that males become better calibrated in their confidence with such practice (Pallier, 2003). It may also be helpful for teachers to go over wrong answers with high confidence ratings and correct answers with low confidence ratings. This shows areas where misunderstandings have occurred and are key places for additional instruction.

Discussions regarding confidence calibration and resolution can be combined with discussions about gender and cultural stereotypes. For example, some research on confidence has found that girls are less likely to take risks, and that high-achieving girls are less confident on unfamiliar tasks (Jones & Jones, 1989). Teachers might encourage girls to take on challenging, unfamiliar tasks and try to disrupt this stereotype. Other research indicates that boys dominate in whole class and in small group settings, even when they do not know the answer (Grossman & Grossman, 1994). In subjects such as mathematics, teachers can be clear about their expectations for small group work, making sure that boys explain their answers to the group and not just ask for help or give answers without explanations. In science, ensure that girls get to do labs rather than just taking notes or engaging in other stereotypical female behaviors.

Finally, in preparing future teachers or in continuing professional development with teachers at either the university, building, or district level, we recommend that teachers look at institutional as well as individual constraints and expectations regarding gender, culture and confidence. The influence of gender and cultural stereotypes is moderated by children's identity (Meece, Glienke, & Burg, 2006), and identity is, in part, shaped by classroom contexts. Girls and boys are not homogeneous groups, so we encourage teachers and researchers to examine how race, socioeconomic status, and country may influence identity development and school experiences.

Conclusion

With regard to gender and cultural differences in confidence, context matters. Some of the contexts that seem to matter involve how confidence is examined, what discipline is tested, characteristics of students, the culture of students, and the context in which the research was conducted. Gendered behaviors and cultural behaviors are, to some extent, socially constructed, that is, both situated in specific contexts and constructed through social interaction (Francis & Skelton, 2005). Thus, when we discuss gender or cultural

differences, we have to consider the above contexts, which are likely to have influenced the research outcomes.

Gendered behaviors and cultural behaviors are also influenced by discourse and by time. For example, whereas researchers used to lament women's "underachievement" in mathematics and science, a meta-analysis of performance shows that with regard to quantitative skills, women are gaining faster than the gap can travel (Rosenthal & Rubin, 1982). The gender gap in motivation in science and mathematics and achievement narrows with age and time, whereas the gap in literacy remains throughout the school years (Meece et al., 2006). In England and Wales, after the introduction of the National Curriculum, which requires boys and girls to enroll in similar subjects, girls caught up with boys in math and science, whereas boys did not catch up with girls in English and modern languages. This "underachievement" of boys, however, has not hampered their careers in the same way girls were previously constrained (and continue to be) in careers involving mathematics and science (Francis & Skelton, 2005). As we have said previously, with regard to what was previously considered females' lack of confidence, may be a problem with discourse and with using males as a standard for performance (Lundeberg et al., 1994). We have found that women are often as confident as men when they know something, but when they are unsure of their knowledge, they admit uncertainty. Some men, however, do not seem to be as good at knowing when they know something and when they do not. Thus, more appropriate discourse might be men's overconfidence, rather than women's lack of confidence. However, this finding is also moderated by time.

In this chapter, we have focused on the role of confidence in academic settings; however, calibrating confidence is also important in life outside the classroom. For example, in diving, women are more accurate evaluators of their performance than are men and respond better to criticism from coaches (Feltz, 1988). In business, men are overconfident in stock market investments and this overconfidence results in money lost—these authors found significant gender differences in money gained via investments for women as compared to men (Barber & Odean, 2001). As the humorist Josh Billings wrote, "It ain't what a man doesn't know that makes him a fool; it is what he does know that ain't so." However, whether males make themselves "a fool" is also dependent on culture and context, because these variables affect both confidence calibration and confidence discrimination.

References

- American Psychological Association (2007). Retrieved August 1, 2007 from <http://apastyle.apa.org/sexuality.html>
- Barber, B. M. & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 261–292.
- Beyer, S. (1999). Gender differences in the accuracy of grade expectancies and evaluations. *Sex Roles: A Journal of Research*, 41(3/4), 279–296.
- Boldizar, J. P. (1991). Assessing sex typing and androgyny in children: The Children's Sex Role Inventory. *Developmental Psychology*, 27(3), 505–515.
- Britner, S. & Pajares, F. (2001). Self-efficacy beliefs, motivation, race, and gender in middle school science. *Journal of Women and Minorities in Science and Engineering*, 7(4), 269–283.
- Burton, S. (Ed.) (1967). *The home book of quotations: Classical and modern* (10th ed.). New York: Dodd, Mead, & Company.
- Casey, M., Nuttall, R., & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education*, 32(1), 28–57.

- Ercikan, K., McCreith, T., & Lapointe, V. (2005). Factors associated with mathematics achievement and participation in advanced mathematics courses: An examination of gender differences from an international perspective. *School Science and Mathematics, 105*(1), 5–14.
- Feltz, D. (1988). Gender difference in the casual elements of self-efficacy on a high avoidance motor task. *Journal of Sport and Exercise Psychology, 10*, 151–166.
- Francis, B. & Skelton, C. (2005). *Reassessinig gender and achievement*. New York: Routledge.
- Glenberg, A. M. & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15*, 84–93.
- Graham, S. & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools—A Report to the Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.
- Grossman, H. & Grossman, S. (1994). *Gender issues in education*. Needham Heights: Allyn and Bacon.
- Harter, S., Waters, P., & Whitesell, N. (1997). Lack of voice as a manifestation of false self-behaviour among adolescents: The school setting as a stage upon which the drama of authenticity is enacted. *Educational Psychologist, 32*, 153–173.
- Heine, S., Kitayama, S., & Lehman, D. (2001). Cultural differences in self-evaluation: Japanese readily accept negative self-relevant information. *Journal of Cross-Cultural Psychology, 32*(4), 434–443.
- Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14*, 299–324.
- Jones, L. P. & Jones, L. G. (1989). Context, confidence and the able girl. *Educational Research, 31*, 189–194.
- Jonsson, A. & Allwood, C. (2001). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences, 34*, 559–574.
- Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*, 582–591.
- Kim, B. & Omizo, M. (2005). Asian and European American cultural values, collective self-esteem, acculturative stress, cognitive flexibility, and general self-efficacy among Asian American college students. *Journal of Counseling Psychology, 52*(3), 412–419.
- Klassen, R. (2004). A cross-cultural investigation of the efficacy beliefs of south Asian immigrant and Anglo Canadian nonimmigrant early adolescents. *Journal of Educational Psychology, 96*(4), 731–742.
- Kleitman, S. & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. *Journal of Applied Cognitive Psychology, 15*, 321–341.
- Koivula, N., Hassmen, P., & Hunt, D. (2001). Performance on the Swedish scholastic aptitude test: Effects of self-assessment and gender. *Sex Roles: A Journal of Research, 44*(11/12), 629–637.
- Krauthammer, C. (1990, February 5). Education: Doing bad and feeling good. *Time, 135*(6), 78.
- Lieberman, V. & Tversky, A. (1993). On the evaluation of probability judgments: Calibration, resolution, and monotonicity. *Psychological Bulletin, 114*, 162–173.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of art. In H. Jungerman & G. DeZeeuw (Eds.), *Decision making and change in human affairs*. Dordrecht, Holland: Reidel.
- Lloyd, J., Walsh, J., & Yilagh, M. (2005). Sex differences in performance attributions, self-efficacy, and achievement in mathematics: If I'm so smart, why don't I know it? *Canadian Journal of Education, 28*(3), 384–408.
- Lundeberg, M. A. & Fox, P. W. (1991). Do laboratory findings on text expectancy generalize to classroom outcomes? *Review of Educational Research, 61*(1), 94–106.
- Lundeberg, M. A., Fox, P. W., Brown, A. C., & Elbedour, S. (2000). Cultural influences on confidence: Country and gender. *Journal of Educational Psychology, 92*(1), 152–159.
- Lundeberg, M. A., Fox, P. W., & Puncochar, J. (1994). Highly confident, but wrong: Gender

- differences and similarities in confidence judgments. *Journal of Educational Psychology*, 86 (1), 114–121.
- McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal of Research in Mathematics Education*, 37(2), 129–150.
- Meece, J. L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. *Journal of School Psychology*, 44, 351–373.
- Pajares, F. (1996). Self-Efficacy Beliefs and Mathematical Problem-Solving of Gifted Students. *Contemporary Education Psychology*, 21, 325–344.
- Pajares, F. & Valiante, G. (2001). Gender differences in writing motivation and achievement of middle school students: A function of gender orientation? *Contemporary Educational Psychology*, 26, 366–381.
- Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles: A Journal of Research*, 48(5/6), 265–276.
- PISA (2003). Retrieved June 15, 2007 from http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_11,00.html
- Pressley, M. & Ghatala, E. S. (1989). Metacognitive benefits of taking a test for children and young adolescents. *Journal of Experimental Child Psychology*, 47, 430–450.
- Pressley, M., Levin, J. R., Ghatala, E. S., & Ahmad, M. (1987). Test monitoring in young grade school children. *Journal of Experimental Child Psychology*, 43, 96–111.
- Puncochar, J. M. & Fox, P. W. (2004). Confidence in individual and group decision making: When “two heads” are worse than one. *Journal of Educational Psychology*, 96(3), 582–591.
- Puncochar, J., Fox, P. W., Fritz, W., & Elbedour, S. (1996, April). Highly confident, but wrong, revisited overseas. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Rosenthal, R. & Rubin, D. B. (1982). Meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology*, 74(5), 708–712.
- Sabina, K. & Lazar, S. (2005). Self-confidence and metacognitive processes. *Zbornik Instituta za pedagoska istrazivanja*, 37(1), 45–69.
- Saunders, J., Davis, L., Williams, T., & Williams, J. (2004). Gender differences in self-perceptions and academic outcomes: A study of African American high school students. *Journal of Youth and Adolescence*, 33(1), 81–90.
- Shaughnessy, J. J. (1979). Confidence-judgment accuracy as a predictor of test performance. *Journal of Research in Personality*, 13, 505–514.
- Stankov, L. & Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences*, 21(6), 971–986.
- Stetsenko, A., Little, T., Gordeeva, T., Grasshof, M., & Oettingen, G. (2000). Gender effects in children’s beliefs about school performance: A cross-cultural study. *Child Development*, 71(2), 517–527.
- Stevens, T., Olivárez, Jr., A., & Hamman, D. (2006). The role of cognition, motivation, and emotion in explaining the mathematics achievement gap between Hispanic and white students. *Hispanic Journal of Behavioral Science*, 28(2), 161–186.
- Troia, G. A. (2007, July) Informative and persuasive text composing strategies for struggling writers. Paper presented at the Literacy Institute 2007, Washington, DC.
- Wright, G. N. & Phillips, L. D. (1980). Cultural variations in probabilistic thinking: Alternative ways of dealing with uncertainty. *International Journal of Psychology*, 15, 239–257.
- Wright, G. N., Phillips, L. D., Whalley, P. C., Choo, G. T., Ng, K., Tan, I., & Wisudha, A. (1978). Cultural differences in probabilistic thinking. *Journal of Cross-Cultural Psychology*, 9, 285–299.
- Yanniv, I., Yates, J. F., & Smith, J. E. S. (1991). Measures of discrimination skill and probabilistic judgment. *Psychological Bulletin*, 110, 611–617.
- Yates, F., Lee, J., & Shinotsuka, H. (1996). Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes*, 65(2), 138–147.
- Yates, J. F., Zhu, Y., Ronis, D. L., Wang, D., Shinotsuka, H., & Toda, M. (1989). Probability

judgment accuracy: China, Japan, and the United States. *Organizational Behavior and Human Decision Process*, 43, 145–171.

Zukerman, H. (1987). Persistence and change in the careers of men and women scientists and engineers: A review of current research. In L. S. Dix (Ed.), *Women: Their underrepresentation and career differentials in science and engineering* (pp. 123–156). Proceedings of a workshop at the National Academy of Sciences. Washington, DC: National Academy Press.

13 Teachers as Metacognitive Professionals

Gerald G. Duffy and Samuel Miller
University of North Carolina-Greensboro

Seth Parsons
George Mason University

Michael Meloth
East Carolina University

This chapter reviews the research on teachers' metacognitive thought. Three questions drive the inquiry. First, "How is metacognition related to teaching?" Second, "Can teachers learn to be metacognitive while engaged in the very complex and multi-dimensional act of teaching?" Finally, "What should be the future research agenda on teachers and metacognition?"

While our charge was to review the research on teacher metacognition, we should make clear at the outset three issues that influence our approach to this task. First, research on teachers as metacognitive professionals is a work in progress. While researchers and educators claim frequently that teachers are metacognitive, detailed characterizations based on empirical qualitative or quantitative evidence are scarce.

Second, we assume that instructional effectiveness, as measured by student achievement, is tied, in part, to teacher metacognition. If it were not, there would be little need to understand what teacher metacognition is or how we might help teachers teach metacognitively.

Third, as teacher educators, we are particularly interested in how to develop teacher metacognitive thought. We know that it does not simply happen. For it to occur, teachers need to make conscious and deliberate decisions when planning and when working with students. Little is known, however, about how to develop this ability. Consequently, our goal for this chapter is to use available research to clarify what is known and unknown about teachers as metacognitive professionals, and to propose a research agenda that advances our understanding both of teachers' metacognitive actions and of how to promote such instructional actions.

Background

Historically, classroom teaching has been described in procedural terms. Certain routines and behaviors are repeated over and over. Based on the results of process-product research from the 1970s (Brophy & Good, 1986), we know that certain routines, such as well-managed lessons that include introductions, detailed description of content, teacher-directed questions, and independent practice, are strongly associated with students' achievement. Such routines can be considered the "What works?" view of teaching. The drawback of this view is that it assumes that improved classroom teaching is limited to procedural methods and techniques when, in reality, classrooms produce unpredictable situations requiring more than established procedures.

While the role of routines in teaching is reasonable up to a point, research conducted since the 1970s has revealed much more sophisticated dimensions of effective teaching (Brophy, 1998). Carter (1990) describes teachers and teaching using terms such as “thoughtfulness,” “decision-making,” and “thinking on one’s feet.” That is, close examination of teachers in classrooms suggests that while the best teachers do employ routines and procedures “that work,” they also engage in complex mental activity that helps them decide how to alter routines and procedures when necessary. Bransford, Darling-Hammond, and LePage (2005) may have expressed it best:

... teaching looks simple from the perspective of students who see a person talking and listening, handing out papers, and giving assignments. Invisible ... are the many kinds of knowledge, unseen plans, and backstage moves—the skunkworks, if you will—that allow a teacher to purposefully move a group of students from one set of understandings and skills to quite another over the space of many months. (p. 1)

Further, research on teaching has illustrated that these less visible aspects of teaching are not easily accomplished. That is, classroom life often involves dilemma-like situations that are not easily resolved (Windschitl, 2002). Buchmann (1990) described it as follows:

Teaching ... routinely involves multiple pairs of obligations—binding claims on one’s attention in which if one is satisfied, the other cannot be ... Such multiplying moral dilemmas are “resolved” in an interconnected series of imperfect decisions that bypass one horn of each dilemma and deal with any residues in the network later. (p. 5)

In short, as Spillane, Reiser, and Reimer (2002) point out, the work of teachers “involves unpredictable human relations not reducible to programmatic routines” (p.390).

Because of these recent insights, teachers—particularly effective teachers—are often described in “metacognitive” terms. For instance, Lin, Schwartz, and Hatano (2005) say effective teachers possess “adaptive metacognition.” Others use similar terms: Bransford, Derry, Berliner, Hammerness, and Beckett (2005) describe what effective teachers do as “adaptive expertise”; Galda and Guice (1997) call it “response-based instruction”; Shulman (1998) describes it as the ability to “. . . transform, adapt, merge, and synthesize, criticize and invent . . .” (p. 519); Duffy (2005) calls it “thoughtfully adaptive teaching”; Lin (2001) calls it “reflective adaptation,” and Little et al. (2007) refer to it as “wise improvisation.” All of the above describe teachers as effective in large measure because they frequently and deliberately engage in conscious, mindful action (or, as we argue, in metacognitive thought) as well as technical or procedural routines. Our task in this chapter is to investigate the extent to which this assumption is true.

Question #1: How is Metacognition Related to Teaching?

Metacognition usually emphasizes “thinking about one’s thinking” and regulation of that thinking (Flavell, 1976, 1979). According to Baker and Beall (2009), metacognition includes self-regulation, strategic control and the self-correction of actions. Applied to students, metacognition is often described in terms of learning new content or concepts or acquiring and refining the use of learning strategies (Hacker, 1998).

It is assumed that teachers’ metacognition, however, is much more complex (Zohar, 2006). While teachers, like students, need to monitor and regulate their cognitive activity and must be strategic when they attempt to solve a problem, teachers have the additional tasks of promoting content learning, identifying appropriate strategies, making moment-

to-moment decisions to ensure students' learning, adjusting for individual differences, and much more.

In answering the question about how metacognition is related to teaching, we (1) describe problems that inhibit our understanding in this area and (2) describe available evidence regarding the extent to which teachers actually engage in metacognitive actions when teaching.

Problems Inhibiting our Understanding of Teacher Metacognition

Three problems inhibit our understanding of the metacognitive aspects of classroom teaching and the extent to which teaching and metacognition are related. The first involves nomenclature; the second involves the situational nature of teaching; and the third involves the difficulties in "getting inside teachers' heads" to gather data to substantiate that teachers are metacognitive.

Nomenclature Problems in Describing Metacognitive Aspects of Teaching

Primary among the labeling difficulties is the fact that terms such as "metacognition" and "metacognitive actions," while firmly established in the metacognitive literature, often are given different labels in the literature on teaching and teachers. For instance, many educators prefer the term "self-regulation" (see, for instance, Perry, Phillips & Dowler, 2004; Winne & Perry, 2000; Zimmerman, 2000; Zimmerman & Schunk, 2001). Paris and Paris (2001), for instance, define self-regulation as "autonomy and control by the individual who monitors, directs, and regulates actions toward goals of information access, expanding expertise, and self-improvement" (p. 89). Such descriptors are very reminiscent of "metacognition," so it is often difficult to discern a substantive difference between "metacognition" and "self-regulation." While it may be that studies of metacognition morphed into self-regulation because researchers realized how metacognition was tied to factors such as strategy use, study skills, expectations, and so on, the two labels nonetheless are often used interchangeably.

Similarly, other educators associate "metacognition" with "reflection." Given that metacognition is "thinking of one's thinking," it is a short step to associating metacognition with "reflecting" on one's thinking (Baker, 2002). Risko, Roskos and Vukelich (2005), for instance, suggest that reflection and metacognition "... are overlapping constructs ... involving deliberate, evaluative, and constructive activity." They point out that metacognition's distinctive characteristics of self-monitoring and self-regulation require that teachers ask themselves questions about their understanding. Hence, they are thinking about their thinking or, as they say, developing a "self-analytic turn of mind."

Likewise, others focus on "teacher as researcher" (Cochran-Smith & Lytle, 1993). This research also resonates with the concept of metacognition because it is assumed that by studying their own work teachers will be more thoughtful. For instance, Baumann and Duffy-Hester (2000) emphasize teacher research because "it involves reflecting on one's teaching and practice, inquiring about it, exploring it, and then taking action to improve or alter it" (p.78). Similarly, Cochran-Smith and Lytle (1993) argue for teacher research because it promotes a deliberative view of teaching in which teachers "use their knowledge to construct perspectives, choose actions, manage dilemmas, interpret and create curricula, make strategic choices, and to a large extent define their own teaching responsibilities" (p. 100). Again, because of the emphasis on thinking about one's thinking, this view resonates with the idea of metacognitive teaching.

In sum, it is sometimes difficult to synthesize the literature on teachers as metacognitive

professionals because so many different terms are used to describe the thoughtful and intentional mental activity in which teachers presumably engage.

Situational Problems in Using Metacognition to Describe Teachers and Teaching

The situative nature of teaching is a second problem in using metacognition to describe teachers and teaching. Three examples are illustrative.

First, because of the complexity of managing 25 or 30 children across a six-hour school day, efficient management of people and time is essential. Hence, in classrooms there is a definite tension between smooth efficiency and thoughtful innovation (Bransford, Darling-Hammond, et al., 2005). The key to efficiency is routine, primarily routines to manage student behavior. Because of the fundamental importance of routines, teachers may assume they cannot afford to spend too much time being intentionally metacognitive, and the metacognitive action that does take place may be limited to situations that threaten to disrupt activity flow, such as how to handle disruptive students. In short, most effective classroom teachers opt for well-planned school days in which efficiency dominates (see, for instance, Brophy, 1998; Brophy & Good, 1986; Wharton-McDonald, Pressley, & Hampston, 1998). Innovation (or metacognitive thinking) is assumed to be restricted to relatively limited situations, such as unexpected student responses and how to adjust interactive instruction to accommodate students who are struggling to make sense of lesson content. In that sense, metacognitive action is situational.

Second, metacognition in teaching is situational as a function of career level. Because classroom life is complex, teachers always strive to simplify and routinize as much as possible, thereby minimizing the amount of effort needed (Evertson, Emmer, & Worsham, 2005). Hence, first-year teachers must engage in high levels of metacognitive thought virtually all the time because everything is new. As they mature as teachers and gain greater expertise, however, they routinize much of what initially required metacognitive thought. The hard work involved in metacognitive thinking is reserved for limited kinds of non-routine situations (Hammerness et al., 2005).

Third, metacognition in teaching is situational because teacher metacognition is subject to contextual variables. For instance, our experience working in schools suggests that some administrators interpret current No Child Left Behind legislation to mean that teachers should only engage in minimal levels of thought (see, for instance, Valli & Buese, 2007), a trend toward procedural and less adaptive teaching that is enforced by “local leaders who demand fidelity to program design” (Correnti & Rowan, 2007, p. 298). That is, teachers need only follow a script and provide lots of drill and practice in order to get students to score well on state tests, a phenomenon Pearson (2007) calls the “McDonaldization of teaching” (p. 154) because teachers, like fast food workers, are being trained to follow a routine. In reality, however, the No Child Left Behind legislation actually calls for “measures that assess higher-order thinking skills and understanding” (Wood, Darling-Hammond, Neill, & Roschewski, May 7, 2007). Such a requirement would compel teachers to engage in frequent metacognitive thought. If legislation is interpreted as technical prescriptions to be applied exactly, the emphasis is on getting students to answer routine questions accurately, and teacher metacognitive actions are discouraged; if, in contrast, legislation is interpreted more liberally, students are encouraged to take more control of their learning, students’ metacognitive thinking is promoted, and teachers’ metacognitive actions are encouraged. So how legislation is interpreted affects the extent to which teachers engage in metacognitive action.

Consequently, there is a situational dimension to the problem of teachers as metacognitive

professionals. How much thought and control a teacher exerts depends on what the teacher is required to do; how frequently the teacher has to do it; students' developmental level, needs, and interests; and the teacher's instructional goals.

Methodological Problems in Substantiating the Nature of Teachers' Metacognition

Teacher thinking has long been a research focus in education (see, for instance, Clark & Peterson, 1986). Similarly, teacher knowledge and teachers' use of knowledge have received much attention (Munby, Russell, & Martin, 2001). Despite all this effort, however, little progress has been made. The bottom line is that it is very difficult to access teacher thinking, and to document the extent to which teachers are metacognitive.

The favored methodological approaches are various forms of stimulated recall and self-reports of thinking. In stimulated recall, researchers and teachers typically view together a videotape of a recent lesson taught by the teacher, with the teacher using that viewing as a stimulus to recall the thinking done during the lesson. In self-report studies, teachers are asked to describe what they learned or what they were thinking during professional development workshops or during classroom instruction. However, it has long been recognized that these methodological approaches are far from perfect (Ericsson & Simon, 1993; Borko & Putnam, 1996). For instance, while teachers often report that professional development designed to help them act more thoughtfully (or metacognitively) *did* cause them to think differently, this effect is seldom observed in their instruction (Tobin, 1993; Watts, Jofili, & Bezerra, 1997). Further, there is the possibility that expert teachers' knowledge becomes so intuitive that it cannot be accessed, a phenomenon Bransford, Stevens, et al. (2006) call "expert blind spots." Hence, the results derived from these methodologies have been limited.

The methodological issue is compounded by the fact that metacognition in teaching is not simply a matter of knowledge or cognition. There is also a dispositional aspect to being metacognitive. That is, to be an effective teacher, one must be disposed to being thoughtfully adaptive in response to complex and unanticipated problems that arise (Meloth & Deering, 1999). Studying teacher metacognition, then, must include emotions as well as cognitions (Dole & Sinatra, 1998; Gregoire, 2003; Pintrich, Marx, & Boyle, 1993; Paris & Winograd, 1990). But teacher emotions are just as difficult to study as teacher cognitions.

Summary

In sum, then, there are several problems associated with the study of teachers as metacognitive professionals. The first is the problem of terminology and the fact that a variety of terms are used to describe teacher thought. The second is the fact that metacognitive action by teachers is situational, varying as a function of setting, students, and career level. The third is that we have few methodological tools for studying the essentially invisible act of teacher thinking.

To What Extent do Teachers Engage in Metacognitive Actions?

Studies of effective teachers would suggest that teachers do engage in metacognitive actions. Such studies emphasize that teachers daily confront unanticipated situations requiring immediate decisions. Consequently, Carter (1990) describes effective teachers as operating from complex and richly elaborated knowledge structures; Berliner (1994)

says effective teachers engage in sophisticated planning and capitalize on varied kinds of instructional opportunities; Roehler, Duffy and Warren (1988) report that effective teachers often make independent, “against the grain” decisions when they feel it is necessary; and Wharton-McDonald et al. (1998) note that effective teachers orchestrate together a wide variety of techniques and strategies that fit a situation but may not be theoretically pure. As the National Reading Panel (NICHD, 2000) stated regarding the teaching of reading comprehension, researchers “have not identified a specific set of instructional procedures that teachers can follow routinely. Indeed, they have found that reading comprehension cannot be routinized” (pp. 4–125). Instead, as Snow, Burns, and Griffin (1998) say, what distinguishes effective teachers is their ability “to craft a special mix of instructional ingredients for every child they work with” (p. 2). Garrison (1997) sums it up by saying that effective teaching “is doing the right thing in the right way and at the right time in response to problems posed by particular people in particular places on particular occasions” (p. 271). In short, nothing works all the time or, as Mosenthal, Lipson, Torncello, Russ, and Mekkelsen (2004) report, no single approach or model accounts for success in schools where large numbers of students are successful (see also, Duffy & Hoffman, 1999).

Studies of exemplary teachers substantiate these findings. For instance, Pressley (2002) states that although exemplary fourth-grade teachers plan their instruction well, “they also take advantage of teachable moments by providing many apt mini-lessons in response to student needs” (p. xiii). Morrow, Tracey, Woo and Pressley (1999, p. 467) similarly describe exemplary primary-grade teachers as providing “spontaneous skill development.” Duffy, Roehler, Sivan, et al. (1987, p. 364) report that “the best explainers generated spontaneous explanations throughout the lesson, elaborating in response to students’ restructured understandings . . .” and Taylor, Pearson, Clark, and Walpole (2000) reported “on-the-fly” instruction to be characteristic of their exemplary teachers. Wharton-McDonald et al. (1998) sum it up well when they state:

Although the high achievement teachers followed predictable patterns of activities and expectations, they were not rigid in their adherence to planned lessons. Within the overall structure of a lesson, they maintained the flexibility to pursue topics that arose in discussion, or to insert mini-lessons when the need became apparent. (p. 120)

Similarly, studies of expert teachers, while not focusing specifically on metacognition, often provide characterizations closely aligned with metacognition. Berliner (2004), for example, offers the following description of expert teachers:

. . . expert teachers are more sensitive to the task demands and social situation when solving pedagogical problems; expert teachers are more opportunistic and flexible in their teaching than are novices; expert teachers represent problems in qualitatively different ways than do novices; expert teachers have fast and accurate pattern-recognition capabilities, whereas novices cannot always make sense of what they experience; expert teachers perceive meaningful patterns in the domain in which they are experienced; and although expert teachers may begin to solve problems slower, they bring richer and more personal sources of information to bear on the problem that they are trying to solve. (p. 201)

This profile suggests that expert teachers are metacognitive because they make generalizations and construct rules to guide their actions (Kuhn, 2001), are aware (or can easily bring to a conscious level) an action that is likely to be successful and, more importantly,

why it should be successful (Zohar, 2006), and draw as much upon their knowledge of subject matter as their knowledge of the learner (Grimmett & MacKinnon, 1992).

On the surface, then, the best teachers have proactive states of mind and emotional strength which allows them to “take charge” of the complexities and uncertainties of daily teaching. The assumption is that they are metacognitive. That is, they make adaptive decisions as they teach because the unpredictability of the classroom and the nature of students’ learning means that teaching can never be completely routinized. But none of the above reports were studying teachers’ metacognitive thought per se. Instead, they were making more general observations of teachers and, on the basis of those observations, concluding that these teachers *must have been* metacognitive. But do we have empirical evidence to document this seemingly apparent characteristic?

One logical place to look is the research on teacher decision-making. Historically, teachers have been described as making thousands of decisions daily (Doyle, 1977; Jackson, 1968). Since 1975, when the Institute for Research on Teaching at Michigan State University initiated research on teachers as decision-makers, numerous researchers have conducted studies in this area (see, for instance, Clark & Peterson, 1986; Englert & Semmel, 1983). Shavelson (1973) has even declared that decision-making is the primary teaching skill. It has been assumed that such decision-making is metacognitive in nature and, in fact, much of the case for teachers as metacognitive professionals is rooted in the assumption that teachers make deliberate decisions, or pedagogical moves. However, the evidence in this area is thin.

Another logical place to look for evidence of teachers being metacognitive is the research on scaffolding. Scaffolding is that part of lessons where teachers provide gradually diminishing amounts of assistance as students become confident in learning the task. Pearson and Gallagher (1983) have called this “gradual release of responsibility,” Duffy and Roehler (1987) have called it “responsive elaboration,” and Pressley and his colleagues (1992) have called it “transaction.” Researchers of instructional scaffolding, such as Maloch (2002), Many (2002) and Rodgers (2004/5) report that during scaffolding teachers follow up with questions, clarify, provide cues and elaborate to meet diverse student needs. It is assumed that such actions are metacognitive in that they require teachers to regulate and control knowledge in deciding when and how to insert these spontaneous scaffolds. As with decision-making research, however, we have no empirical data describing and categorizing such thoughtful actions. And we do not have research that ties such actions to improved student achievement.

However, a third area of research is now attempting to provide such data (Duffy, Miller, Kear, Parsons, Davis, & Williams, 2007). Utilizing collective case studies, this research examines the types and quality of elementary literacy teachers’ thoughtfully adaptive actions during instruction and the types and quality of their rationales for these adaptations. To date, results from the first 27 cases suggest that teachers make relatively few high-quality adaptations during instruction and provide relatively few high-quality rationales for those adaptations.

The most recent study in this line of research suggests that the relative lack of metacognitive thought may be tied to the tasks teachers implemented (Parsons, 2008). This study indicated that third-grade literacy teachers’ adaptations were closely tied to the openness of students’ tasks. That is, when teachers implemented authentic, challenging tasks, they adapted their instruction in more thoughtful ways than when they implemented closed tasks such as worksheets. Hence, teachers’ metacognitive thought may be related to the type of instruction they implement.

While it is interesting to consider that teacher thoughtfulness may be associated with type of instruction, ensuring open tasks in the current context of No Child Left Behind

legislation is problematic. Evidence provided earlier in this chapter (see, for instance, Duffy, Roehler, & Putnam, 1987; Duffy & Anderson, 1984; Duffy & Ball, 1986; Valli & Buese, 2007) consistently points to the fact that teachers' decisions are not regulated by teacher thought as much as by contextual conditions such as centrally imposed directives and local conditions in the classroom environment. For instance, Grossman, Valencia, et al. (2000) reported that "aspects of school and district context . . . can support or thwart continuing learning and fuller appropriation of ideas and practices" (p. 660) while Maloch et al. (2003) report "a sense of disempowerment" among today's teachers who view their role as "one of compliance" (p. 446).

In sum, our question about whether teachers are metacognitive professionals cannot be definitively answered. Observations of exemplary teachers strongly suggest that effective teachers regulate and control their thinking as they teach. There is hardly any other way to explain what researchers see good teachers doing. However, studies designed to examine teachers' metacognitive thought, such as those cited earlier, suggest that teachers in general may not rise to the level of metacognitive thought presumably associated with exemplary teachers (although, as noted, current policy issues that promote procedural teaching as opposed to thoughtful teaching may influence those situations). In sum, while recent research on thoughtfully adaptive teaching is beginning to yield data regarding the nature of teacher metacognition and the circumstances under which it occurs (Parsons, 2008), this line of research is in its infancy. Most important of all, of course, we have no empirical evidence that teacher metacognition is associated with improved student learning. In short, it is assumed that teachers are metacognitive, but more data is needed to document the extent to which they are metacognitive, the factors influencing it, and the effect on students.

Question #2: Can Teachers Learn to be Consistently Metacognitive?

Because metacognitive thought is frequently mentioned as a desirable teacher characteristic, professional developers and teacher educators are exploring how to develop it. Consistent with past research on professional development indicating that educative approaches increase teacher thoughtfulness (Richardson & Placier, 2001), we focus on an educative model. In contrast to traditional professional development that emphasizes technical implementation, short training periods, and authority figures as leaders, educative models of professional development have three starkly different characteristics. First, educative professional development is dynamic in that teachers are given control, are encouraged to take positions and to debate them, and are encouraged to use professional knowledge as a basis for constructing their own understandings of how best to improve practice (Anders & Richardson, 1991). Second, the development is long term. That is, it occurs over months and years, not over days or weeks (Duffy, 2003). Third, they are situated, in that they are "case-based" or "problem-based" around actual situations teachers are facing (Shulman & Shulman, 2004) and involve teachers as equal participants in learning communities described as "collaborative innovation" (Randi & Corno, 2000) or "collaborative problem solving" (Hawley & Valli, 1999).

However, a prevalent characteristic of educative professional development is that it is often met with teacher resistance (see, for instance, Duffy, 1993). There are two major reasons for this. First, educative professional development usually emphasizes teacher thoughtfulness, and teachers often resist being thoughtful. They do so because experience has taught them that it is safer and easier to operate from routines in the current policy environment, and because their "apprenticeship of observation" (Lortie, 1975) as students themselves for 13 years causes them to think that they already know how to teach (Kanfer & Kanfer, 1991; Kennedy, 1999). The result is often what Windschitl (2002)

called “additive” change or what Huberman (1990) called “tinkering,” in which teachers insert minor changes into their existing practices. Second, research indicates that learning to be thoughtful occurs in erratic spurts and not as steady growth. For instance, in teaching teachers how to implement strategy instruction in thoughtful ways over the course of a four-year professional development project, Duffy (1993) noted that teachers proceeded in an erratic fashion through a series of eight “points of progress” that at times represented steady growth but at other times slowed or stopped.

In an attempt to overcome these difficulties, professional developers have initiated two kinds of instructional emphases: a focus on teacher knowledge and a focus on developing the disposition or propensity to be thoughtful.

From the perspective of teacher knowledge, many educators have described the kinds of knowledge teachers presumably need to be effective (Carter, 1990; Bransford, Darling-Hammond, & LePage, 2005; Fenstermacher, 1994; Munby, Russell, & Martin, 2001). However, there has been no clear agreement. For instance, a recent debate pits emphasis on declarative (i.e., knowing that) and procedural knowledge (i.e., knowing what) against thoughtfulness, with one side arguing that teachers should first learn declarative and procedural knowledge while delaying analysis and reflection until later in the inservice years (Snow, Griffin, & Burns, 2005) and with the other arguing that reasoned analysis should be emphasized from the very beginning (Hammerness et al., 2005).

Two recent emphases take a different approach, however. One examines the role of teachers’ “personal” or “practical” knowledge (see, for instance, Elbaz, 1983). Levin and He (2005) refer to these as “personal practical theories” teachers use in deciding how to proceed pedagogically. They are associated with a metacognitive dimension of teaching because such theories are viewed as tools for reflecting on one’s work.

Another examines teachers from a dispositional perspective. This approach assumes that the propensity to be metacognitive is rooted in a sense of self (Paris, Byrne, & Paris, 2001) or in a volition or will (Randi, 2004). That is, based on self-knowledge, teachers develop a propensity for acting in certain ways. The result is a sort of personal agency based in conscious guides or principles teachers use to sort through the ambiguity and complexity associated with being metacognitive while teaching.

Research on teacher identity falls into this category (Flores & Day, 2006). Identity refers to teachers’ propensity to engage in practices influenced by socially imposed storylines. This sociocultural perspective argues that a teacher’s identity is always in flux and is adapted to circumstances in the context. Nonetheless, identity is assumed to be influential in shaping a teacher’s disposition to act in certain ways.

Similarly, another perspective examines visioning, a sense of self that is presumed to come from the ideal outcomes the teacher seeks to attain (Duffy, 2005; Hammerness, 2003; 2004; Turner, 2006). Vision can focus on what the teacher will ideally become (an egocentric vision) or on what the teacher’s students will ideally become (a student-centered vision). In either case, the vision is metacognitive in the sense that teachers consciously operate from these ideals.

In sum, it is often assumed that teachers can learn to be metacognitive, and that both cognitive and dispositional aspects of professional development can be designed to intentionally encourage teachers to become metacognitive professionals. However, the hard fact is that current efforts are based primarily on common sense and theories rooted in constructivism and sociocultural thinking, and no empirical data exists to substantiate that one or another kind of professional development results in teachers who are metacognitive.

Question #3: What Should be the Future Research Agenda for Teachers and Metacognition?

In this chapter, we have raised several areas of concern. It should be no surprise, then that our research recommendations stem in large part from these concerns. Before we outline the various avenues of research, however, we return to the problem of labels.

We have made the case that teachers' metacognitive actions have a number of labels, and that there are scores of studies on the mental aspects of teaching, all having a metacognitive flavor. While we in no way are advocating that scholars should rename what they study, we would like to suggest that metacognition may provide a theoretical umbrella for these studies and that, as such, researchers should report on the metacognitive elements inherent in their work. Doing so might help all of us identify the metacognitive dimensions of our particular approaches to teacher thinking.

Turning now to future directions, it is apparent that there is only limited empirical support for the widely held assumption that teachers are metacognitive. Initiating an agenda of programmatic research is, therefore, imperative. Given the limited evidence we have cited in this chapter, coupled with anecdotal reports too numerous to list, it is likely that good teachers are metacognitive when they encounter unanticipated problems or unique situations. What we need is more empirical evidence describing teachers' metacognition in all its rich detail so that we know more about how to help future and current teachers teach effectively, and so we can help policy-makers create the kind of educational policies that give teachers room to move instructionally.

One such line of research is already under way. The focus is teachers' thoughtful actions and their metacognitive thought relative to those actions (Duffy et al., 2007). A necessary next step is to note the impact on student achievement. This last aspect is an essential because teacher metacognition is important only to the extent that we can establish that it has a positive impact on student learning. If we wish to influence educational policy in the current political environment, it will be necessary to establish that teachers' metacognitive actions positively impact student performance on state and national tests as well as on measures of higher-order thinking more typically associated with metacognition.

In pursuing such study, we should also seek out expert teachers who work with struggling learners. Studying teachers who work with highly successful students is certainly important as well, but the reasons some students are successful may have less to do with how their teacher teaches and more to do with their entering knowledge and skills. Smart students often do well in classrooms despite their teachers. If we are to understand how teachers' metacognition has an impact on students' learning, we need to focus on classrooms populated with students who struggle.

Once a connection has been established between teachers' metacognitive actions and student achievement, it will be necessary to examine more closely the *way* in which teachers adapt. That is, how do they decide to make the adaptations they do? What knowledge do they access? What mental process do they engage in? Obviously, we must learn more about this process if we are to teach other teachers to engage in metacognitive actions.

Pursuing the above questions about teacher knowledge should alter some of our current understandings. As noted earlier in this chapter, much of the research on teacher knowledge has revolved around the issues of "declarative" and "procedural" knowledge—that is, knowledge "about" things. For instance, Bransford, Darling-Hammond and LePage (2005) recommend that teachers must know about learners, about subject matter and about teaching. But teachers engaging in metacognitive actions seem to be accessing more than just declarative and procedural knowledge. Their thinking seems to go beyond the technical level—beyond "what" and "how" issues that could be tested with multiple-

choice test items. For instance, Pearson (2007) calls for “broad and deep” knowledge in which the hallmarks are “flexibility” and “versatility” in applying knowledge to the differential needs of students. So what kind of knowledge is this? Is it conditional knowledge about the “why” and “when” of content, learners, and pedagogy? From the perspective of how to develop metacognitive professionals, declarative and procedural knowledge seem to encourage “pat” or technically defined answers while emphasis on conditional knowledge may encourage broader, longer-term thinking based in “why” and “when” issues. Or is metacognitive teaching rooted in a still broader form of conceptual knowledge that focuses teachers on long-range outcomes as opposed to short-term goals? Or is it a still broader understanding embedded in what Bransford, Darling-Hammond, and LePage (2005) describe as “a vision of professional practice”? Understanding the mental process involved in a flexible and versatile application of knowledge is crucial if we are to teach teachers to be metacognitive.

Of course, as teacher educators we are particularly interested in how to develop teachers who are metacognitive. Little actual research has focused on whether teachers can learn to be metacognitive and, if so, how to develop them. Two notable exceptions are the work of Perry and her colleagues (2004, 2006) and Parsons (2008). The former worked with preservice teachers in the area of self-regulated learning and asserted that teachers who emphasize open-ended tasks not only promote self-regulated learning in their students but also learn to be self-regulating themselves, while the latter suggests potentially fruitful avenues regarding how to develop such ability. The underlying issue lurking behind all such questions, of course, is how to study a construct such as teacher metacognition. We definitely must devote greater efforts to solving the methodological problems associated with studying teachers’ metacognition. We cannot observe teachers’ metacognition directly because, obviously, it is not possible to directly observe mental phenomena. Thus, despite their shortcomings, we must continue to explore verbal reports of thinking. We conceive of verbal reports broadly—free-ranging interviews about general practice, focused responses while reviewing a video of a particular interaction with a student, personal journals and the like are all viable sources. But whatever the source, these reports must be triangulated with several additional data sources, such as observations of specific teaching actions and records of how teachers modify their actions and their curriculum over time.

Qualitative and quantitative studies of teachers’ metacognition and how it is enacted during instruction are both essential if we are to advance our knowledge. Qualitative studies that link verbal reports of thinking before and during instruction to actions while teaching can be especially advantageous in the clarification of labeling and situational issues of metacognition. These studies can offer detailed insights into how individual teachers think and act.

Regarding quantitative studies, traditional comparative studies will remain necessary because they offer an opportunity to identify differences in, for example, how teachers who operate from a metacognitive stance differ from those operating from a less thoughtful stance. Correlational studies will also remain important because they offer glimpses of the possible relationship among teachers’ thinking, their actions, and student learning.

It is also essential, as noted several times already, that we conduct studies that can more directly connect teacher metacognition to student learning. The causal connection between teacher thinking and student learning is exceptionally complex. Analysis of variance approaches, including multivariate and repeated measures designs, cannot go far enough to establish causal connections because subjects are not fully independent and the methodology does not permit disentangling individual and group effects (Bryk & Raudenbush, 1992). What is needed, then, are “chain of evidence” studies (Fishman, Marx, Best, & Tal,

2003) that involve more sophisticated analysis approaches, such as hierarchical linear modeling (HLM). HLM is advantageous because it is specifically designed to examine how environmental variables (e.g., teacher thinking, teacher actions, curricular differences across classrooms, etc.) affect individual outcomes, such as student achievement. While we value what qualitative studies can yield using single-subject designs, thick descriptive techniques, and collective cases, we are also unapologetic in our belief that statistical data will be necessary because if we cannot ultimately show improvement in children's learning, then the real purpose of studying teachers and teaching is lost.

Conclusion

The field is conflicted regarding the issue of teachers as metacognitive professionals. On the one hand, society's stated desires for what future citizens need argues for thoughtful educational experiences for children which, in turn, require thoughtful, responsive teachers such as those identified as most successful in studies of effective and exemplary teachers. On the other hand, however, there is limited evidence regarding precisely what a metacognitive teacher does, how he or she does it, whether doing it is key to student achievement, or how teacher educators can develop it.

In sum, then, the major message of this chapter is that while we spend a great deal of time promoting the concept of teachers as metacognitive professionals, we are way overdue on conducting the research necessary to substantiate the validity of this idea and to justify the energies being devoted to it. Consequently, we end with a plea for direct scientific studies designed to determine systematically whether teacher metacognition influences student learning in desirable ways and, if it does, how teacher educators and professional developers can develop metacognition in teachers.

References

- Anders, P. & Richardson, V. (1991). Research directions: Staff development that empowers teachers' reflection and enhances instruction. *Language Arts*, 68, 316–321.
- Baker, L. (2002). Metacognition in comprehension instruction. In C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 77–95). New York: Guilford.
- Baker, L. & Beall, L. (2009). Metacognitive processes and reading comprehension. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 373–388). Mahwah, NJ: Lawrence Erlbaum Associates.
- Baumann, J. F. & Duffy-Hester, A. M. (2000). Making sense of classroom worlds: Methodology in teacher research. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 77–98). Mahwah, NJ: Lawrence Erlbaum.
- Berliner, D. C. (1994). Expertise: the wonder of exemplary performances. In J. Mangieri & C. C. Block (Eds.), *Creating powerful thinking in teachers and students; Diverse perspectives* (pp. 161–186). Fort Worth, TX: Harcourt Brace College Publishers.
- Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology & Society*, 24(3), 200–212.
- Borko, H. & Putnam, R. T. (1996). Learning to teach. In D. L. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 673–708). New York, NY: MacMillan.
- Bransford, J., Darling-Hammond, L., & LePage, P. (2005). Introduction. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 1–39). San Francisco, CA: Jossey-Bass.
- Bransford, J., Derry, S., Berliner, D., Hammerness, K., & Beckett, K. L. (2005). Theories of learning and their roles in teaching. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 40–87). San Francisco, CA: Jossey-Bass.

- Bransford, J., Stevens, R., Schwartz, D., Meltzoff, A., Pea, R., Roschelle, J. (2006). Learning theories and education: Toward a decade of synergy. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 209–244). Mahwah, NJ: Lawrence Erlbaum.
- Brophy, J. (1998). Classroom management as socializing students into clearly articulated roles. *Journal of Classroom Interaction*, 33, 1–4.
- Brophy, J. & Good, T. (1986). Teacher behaviors and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). NY: Macmillan.
- Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.
- Buchmann, M. (1990). Beyond the lonely, choosing will: Professional development in teacher thinking. *Teachers College Record*, 91, 481–508.
- Carter, K. (1990). Teachers' knowledge and learning to teach. In W. Houston (Ed.), *Handbook of research on teacher education* (pp. 291–310). NY: MacMillan.
- Clark, C. M. & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of the research on teaching* (3rd ed., pp. 255–296). New York: MacMillan.
- Cochran-Smith, M. & Lytle, S. L. (1993). *Inside/outside: Teacher research and knowledge*. New York: Teachers College.
- Correnti, R. & Rowan, B. (2007). Opening up the black box: Literacy instruction in schools participating in three comprehensive school reform programs. *American Educational Research Journal*, 44, 298–338.
- Dole, J. A. & Sinatra, G. M. (1998). Reconceptualizing change in the cognitive construction of knowledge. *Educational Psychologist*, 32, 109–128.
- Doyle, W. (1977). Learning in the classroom environment: An ecological analysis. *Journal of Teacher Education*, 28, 51–55.
- Duffy, G. G. (1993). Teachers' progress toward becoming expert strategy teachers. *Elementary School Journal*, 94, 109–220.
- Duffy, G. G. (2003). Teachers who improve reading achievement: What research says about what they do and how to develop them. In D. Strickland & M. Kamil (Eds.), *Improving reading achievement through professional development* (pp. 3–22). Norwood, MA: Christopher-Gordon.
- Duffy, G. G. (2005). Developing metacognitive teachers: Visioning and the expert's changing role in teacher education and professional development. In S. E. Israel, C. C. Block, K. L. Bauserman, & K. Kinnucan-Welsch (Eds.), *Metacognition in literacy learning: Theory, assessment, instruction, and professional development* (pp. 299–314). Mahwah, NJ: Lawrence Erlbaum.
- Duffy, G. G. & Anderson, L. (1984). Teachers' theoretical orientations and the real classroom. *Reading Psychology*, 5(1), 97–104.
- Duffy, G. G. & Ball, D. L. (1986). Instructional decision-making and reading teacher effectiveness. In J. V. Hoffman (Ed.), *Effective teaching of reading: Research and practice* (pp. 163–180). Newark, DE: International Reading Association.
- Duffy, G. G. & Hoffman, J. V. (1999). In pursuit of an illusion: The flawed search for a perfect method. *Reading Teacher*, 53(1), 10–17.
- Duffy, G. & Roehler, L. (1987). Improving classroom reading instruction through the use of responsive elaboration. *Reading Teacher*, 40, 514–521.
- Duffy, G., Miller, S. Kear, K. Parsons, S. Davis, S., & Williams, B. (2007, December). *Teachers' instructional adaptations during literacy instruction*. A paper presented at the National Reading Conference. Austin, TX.
- Duffy, G. G., Roehler L. R., & Putnam, J. (1987) Putting the teacher in control: Basal reading textbooks and instructional decision making. *The Elementary School Journal*, 87(3), 357–366.
- Duffy, G., Roehler, L., Sivan, E., Rackliffe, G., Book, C., Meloth, M., Vavrus, L., Wesselman, R., Putnam, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, 22, 347–368.
- Elbaz, F. (1983) *Teacher thinking: A study of practical knowledge*. New York: Nichols Publishing.
- Englert, C. S. & Semmel, M. I. (1983). Spontaneous teacher decision making in interactive instructional contexts. *Journal of Educational Research*, 77(2), 112–21.

- Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised edition). Cambridge, MA: The MIT Press.
- Evertson, C., Emmer, E. T., & Worsham, M. E. (2005). *Classroom management for elementary teachers* (7th ed.). Boston: Allyn & Bacon.
- Fenstermacher, G. D. (1994). The knower and the known: The nature of knowledge in research on teaching. *Review of Research in Education*, 20, 3–56.
- Fishman, B. J., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, 19, 643–658.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231–235). Hillsdale, NJ: Erlbaum.
- Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906–911.
- Flores, M. & Day, C. (2006). Contexts which shape and re-shape new teachers' identities: A multiperspective. *Teaching and Teacher Education*, 22, 219–232.
- Galda, L. & Guice, S. (1997). Response-based reading instruction in the elementary grades. In S. Stahl & D. Hayes (Eds.), *Instructional models in reading* (pp. 311–330). Mahwah, NJ: Lawrence Erlbaum.
- Garrison, J. (1997). *Dewey and Eros: Wisdom and desire in the art of teaching*. New York: Teachers College Press.
- Gregoire, M. (2003). Is it a challenge or a threat? A dual process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review*, 15(2), 147–179.
- Grimmett, P. & MacKinnon, A. (1992). Craft knowledge and the education of teachers. In G. Grant (Ed.), *Review of Research in Education* 18 (pp. 385–456). Washington, DC: The American Educational Research Association.
- Grossman, P. L., Valencia, S. W., Evans, K., Thompson, C., Martin, S., & Place, N. (2000). Transitions into teaching: Learning to teach writing in teacher education and beyond. *Journal of Literacy Research*, 32, 631–662.
- Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1–23). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hammerness, K. (2003). Learning to hope, or hoping to learn: The role of vision in the early professional lives of teachers. *Journal of Teacher Education*, 54, 43–56.
- Hammerness, K. (2004). Teaching with vision: How one teacher negotiates the tension between high ideals and standardized testing. *Teacher Education Quarterly*, 31(4), 33–43.
- Hammerness, K., Darling-Hammond, L., Bransford, J., Berliner, D., Cochran-Smith, M., McDonald, M., & Zeichner, K. (2005). How teachers learn and develop. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 358–389). San Francisco, CA: Jossey-Bass.
- Hawley, W. & Valli, L. (1999). The essentials of effective professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 127–150). San Francisco: Jossey-Bass.
- Huberman, M. (1990). Linkage between researchers and practitioners: A qualitative study. *American Educational Research Journal*, 27, 363–391.
- Jackson, P. (1968). *Life in Classrooms*. NY: Holt, Rinehart & Winston.
- Kanfer, R. & Kanfer, F. (1991). Goals and self-regulation: Application of theory to work settings. In M. Maehr & P. Pintrich (Eds.), *Advances in Motivation and Achievement*, Vol. 7 (pp. 287–326). Greenwich CN: JAI Press.
- Kennedy, M. (1999). The role of preservice teacher education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching a learning profession: Handbook of policy and practice* (pp. 54–85). San Francisco: Jossey-Bass.
- Kuhn, D. (2001). How do people know? *Psychological Science*, 2001, 1–8.
- Levin, B. & He, Y. (2005). Investigating the content and sources of preservice teachers' personal practice theories (PPTs). *Journal of Teacher Education*, 59, 55–68.

- Lin, X. D. (2001). Reflective adaptation of a technology artifact: A case study of classroom change. *Cognition & Instruction, 19*, 395–440.
- Lin, X., Schwartz, D. L., & Hatano, G. (2005). Toward teachers' adaptive metacognition. *Educational Psychologist, 40*(4), 245–255.
- Little, J. W. (Chair), Lampert, M., Graziani, F., Borko, H., Clark, K. K., & Wong, N. (2007, April). *Conceptualizing and investigating the practice of facilitation in content-oriented teacher professional development*. Symposium conducted at the 2007 Annual Meeting of the American Educational Research Association, Chicago, Illinois.
- Lortie, D. L. (1975). *Schoolteacher*. Chicago, IL: University of Chicago.
- McVie, M. (2004). Narrative and the exploration of culture in teachers' discussion of literacy, identity, self, and other. *Teaching and Teacher Education, 20*, 881–899.
- Maloch, B. (2002). Scaffolding student talk: One teacher's role in literature discussion groups. *Reading Research Quarterly, 37*(1), 94–112.
- Maloch, B., Flint, A. S., Eldridge, D., Harmon, J., Loven, R., Fine, J. C., et al. (2003). Understandings, beliefs, and reported decision making of first-year teachers from different reading teacher preparation programs. *Elementary School Journal, 103*, 431–457.
- Many, J. E. (2002). An exhibition and analysis of verbal tapestries: Understanding how scaffolding is woven into the fabric of instructional conversations. *Reading Research Quarterly, 37*(4), 376–407.
- Meloth, M. & Deering, P. (1999). The role of the teacher in promoting cognitive processing during collaborative learning. In A. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 235–256). Mahwah, NJ: Lawrence Erlbaum.
- Morrow, L. M., Tracey, D. H., Woo, D. G., & Pressley, M. (1999). Characteristics of exemplary first-grade literacy instruction. *The Reading Teacher, 52*, 462–476.
- Mosenthal, J., Lipson, M., Torncello, S., Russ, B., & Mekkelsen, J. (2004). Contexts and practices of six schools successful in obtaining reading achievement. *Elementary School Journal, 104*, 343–367.
- Munby, H., Russell, T., & Martin, A. K. (2001). Teachers knowledge and how it develops. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.). Washington DC: AFRA.
- National Institute of Child Health and Human Development. (2000). Report of the National Reading Panel. *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (National Institute of Health pub. no. 00–4769). Washington, DC: National Institute of Child Health and Human Development.
- Paris, S. G., Byrne, J. P., & Paris, A. H. (2001). Constructing theories, identities, and actions of self-regulated learners. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 253–287). Mahwah, NJ: Lawrence Erlbaum.
- Paris, S. G. & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89–101.
- Paris, S. G. & Winograd, P. N. (1990). How metacognition can promote academic learning and instruction. In B. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15–51). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Parsons, S. A. (2008). *Case studies of four teachers: The openness of the tasks they implement, the adaptations they make, and the rationales they offer for adapting*. Unpublished doctoral dissertation, University of North Carolina at Greensboro.
- Pearson, P. D. (2007). An endangered species act for literacy education. *Journal of Literacy Research, 39*, 145–162.
- Pearson, P. D. & Gallagher, M. C. (1983). The instruction of reading comprehension. *Contemporary Educational Psychologist, 8*, 317–344.
- Perry, N. E., Phillips, L., & Dowler, J. (2004). Examining features of tasks and their potential to promote self-regulated learning. *Teachers College Record, 106*(9), 1854–1878.
- Perry, N. E., Phillips, L., & Hutchinson, L. (2006). Mentoring student teachers to support self-regulated learning. *Elementary School Journal, 106*(3), 237–254.

- Pintrich, P. R., Marx, R. W., & Boyle, R. B. (1993). Beyond cold conceptual change: the role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63, 167–199.
- Pressley, M. (2002). Foreword. In R. Allington & P. Johnston (Eds.), *Reading to learn: Lessons from exemplary fourth-grade classrooms* (pp. xi–xvi). New York: Guilford.
- Pressley, M., El-Dinary, P., Gaskins, I., Schudeer, T., Bergman, J., Almasi, L., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *Elementary School Journal*, 92, 511–554.
- Randi, J. (2004). Teachers as self-regulated learners. *Teachers College Record*, 106(9), 1825–1853.
- Randi, J. & Corno, L. (2000). Teacher innovations in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation*. San Diego, CA: Academic Press, pp. 651–685.
- Richardson, V. & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.). Washington, DC: American Educational Research Associations.
- Risko, V. J., Roskos, K., & Vukelich, C. (2005). Reflection and the self-analytic turn of mind: Toward more robust instruction in teacher education. In S. E. Israel, C. C. Block, K. L. Bauserman, & K. Kinnucan-Welsch (Eds.), *Metacognition in literacy learning: Theory, assessment, instruction, and professional development* (pp. 315–333). Mahwah, NJ: Lawrence Erlbaum.
- Rodgers, E. M. (2004/5). Interactions that scaffold reading performance. *Journal of Literacy Research*, 36(4), 501–532.
- Roehler, L., Duffy, G., & Warren, S. (1988). Adaptive explanatory actions associated with teaching of reading strategies. In J. Readance & S. Baldwin (Eds.), *Dialogues in literacy research* (pp. 339–346). 37th Yearbook of the National Reading Conference. Chicago: National Reading Conference.
- Shavelson, R. (1973). What is the basic teaching skill? *Journal of Teacher Education*, 14, 144–151.
- Shulman, L. (1998). Theory, practice and the education of professionals. *Elementary School Journal*, 98, 511–526.
- Shulman, L. & Shulman, J. (2004). How and what teachers learn: A shifting perspective. *Journal of Curriculum Studies*, 36, 257–271.
- Snow, C., Burns, M., & Griffin, P. (Eds.) (1998). *Preventing reading difficulties in young children: A report of the National Research Council*. Washington, DC: National Academy Press.
- Snow, C. E., Griffin, P., & Burns, M. S. (Eds.) (2005). *Knowledge to support the teaching of reading: Preparing teachers for a changing world*. San Francisco: Jossey-Bass.
- Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72, 387–431.
- Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal*, 101, 121–165.
- Tobin, K. (1993). Referents for making sense of science teaching. *International Journal of Science Teaching*, 15, 241–254.
- Turner, J. D. (2006). Preservice teachers' visions of culturally responsive literacy instruction. In J. V. Hoffman, D. L. Schallert, C. M. Fairbanks, J. Worthy, & B. Maloch (Eds.), *55th yearbook of the National Reading Conference* (pp. 309–323). Oak Creek, WI: National Reading Conference.
- Valli, L. & Buese, D. (2007) The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, 44, 519–558.
- Watts, M., Jofili, Z., Bezerra, R. (1997). The case for critical constructivism and critical thinking in science education. *Research in Science Education*, 27, 309–322.
- Wharton-McDonald, R., Pressley, M., & Hampston, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal*, 99, 101–128.
- Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research*, 72, 131–175.

- Winne, P. H. & Perry, N. E. (2000). Measuring self-regulated learning. In P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego, CA: Academic Press.
- Wood, G. Darling-Hammond, L., Neill, M., & Roschewski, P (2007). Refocusing Accountability: Using Local Performance Assessments To Enhance Teaching and Learning for Higher Order Skills. *Briefing Paper Prepared for Members of the Congress of the United States*, May 5, 2007.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation*. San Diego, CA: Academic Press.
- Zimmerman, B. & Schunk, D. (2001). *Self-regulated learning and academic achievement*. NY: Springer-Verlag.
- Zohar, A. (2006). The nature and development of teachers' metastrategic knowledge in the context of teaching higher order thinking. *The Journal of Learning Sciences*, 15(3), 331–377.

Part VII

Self-Regulated Learning

14 Supporting Self-Regulated Learning with Cognitive Tools *

Philip H. Winne and John C. Nesbit
Simon Fraser University

A blue-ribbon team writing for the U.S. Institute of Education Sciences describes a significant challenge that students face:

Psychological research has documented the fact that accurately assessing one's own degree of learning is not something that comes naturally to our species, and fostering this ability is a useful, albeit neglected, component of education . . . Unfortunately, people at all grade levels, from the youngest children through advanced college students, experience failures of metacognition and illusions of knowing in which they think they know something but do not. (Pashler et al., 2007, pp. 1, 24)

Our studies add two further challenges to the challenge of the illusion of knowing. First, quite accomplished learners—university undergraduates—engage study tactics that deviate significantly from what research indicates to be effective approaches to learning (Winne & Jamieson-Noel, 2003). Second, these same relatively sophisticated learners appear quite unaware about how they study (Jamieson-Noel & Winne, 2003; Winne & Jamieson-Noel, 2002).

We agree with the Institute's panel that "students' ability to manage their own studying is one of the more important skills that students need to learn, with consequences that will be felt throughout their lives" (p. 1). But, in light of these empirical findings, a justified yet bleak prediction is that, without support, a very large proportion of students will fail to learn how to learn. In this chapter, we propose approaches that offer grounded hope to address this important need.

A Sketch of Metacognition

Metacognition is a rich construct that blends three major topics in scientific psychology: the nature of knowledge, relations between knowledge and behavior, and agency. While more thorough treatments are available in other chapters of this handbook, we elaborate briefly on these matters to contextualize a case we make later about the potential values of using software systems to support and research forms of metacognition that underlie self-regulated learning (SRL).

* Support for this work was provided by grants to Philip H. Winne from the Social Sciences and Humanities Research Council of Canada (410-2002-1787, 410-2007-1159, and 512-2003-1012), the Canada Research Chair Program and Simon Fraser University; and to John C. Nesbit from the Social Sciences and Humanities Research Council of Canada (410-2007-1157).

One common and useful characterization of knowledge with respect to metacognition is as a production system. According to this model, conditions (or IFs) set a stage for actions (or THENs). For example, if a word is italicized in text its then meaning probably should be committed to memory. The spartan representation of knowledge as a single IF–THEN production or integrated systems of them belies significant complexities. Conditions can be multiple and configured in highly complex arrangements. Information within a single condition can be unreliable and probabilistically diagnostic. Actions can be equally complicated, consisting of multiple and contingent steps that have probabilistic main effects as well as side effects.

In models of human cognition and metacognition, IF–THEN productions model intrinsically non-deterministic links between knowledge and behavior. Specifically, when a learner’s environment, including the internal environment created by cognition, presents or evolves into a particular configuration of conditions—a state—there are five gateways that determine whether a production is realized in behavior (THEN) and, if it can be realized, which particular behavior is realized (Winne, 1982). First, the learner must register the existence of conditions and accurately instantiate their values. Second, the configuration of conditions needs to be validly interpreted as diagnostic of one or one among a set of potential actions. Third, to enact actions, the learner must have capabilities to carry out those actions. Fourth, behavior will be realized only if the learner is motivated to spend resources entailed in behaving. Fifth, the learner needs to have free cognitive capacity and opportunity afforded by the environment to enact chosen actions (Winne, 1992). We summarize these gateways in Table 14.1.

These five gateways to successful engagement in designed instruction elaborate the common two-phase description that partitions metacognitive events into (a) metacognitive monitoring and (b) metacognitive control. In metacognitive monitoring, the learner samples a subset of information in the environment and compares it to a profile of standards. The result of this operation is a profile of differences between the perception and standards. Then, on the basis of that profile of differences as well as qualities of the operations that generated it (e.g., the delay in accessing information) the learner chooses actions.

The fourth gateway that links conditions and actions, where motivation enters the picture, requires swimming in the complexities of views about agency (Martin, 2004). Agency is key in conceptions of metacognition as an underlying model of SRL (Winne, 2006). On the surface, agency seems a straightforward construct, a mode of being wherein a learner can choose whether and how to act in the context afforded by arrays of conditions. At deeper levels, agency proves to be a contingent quality of human action and perhaps one not always under deliberate control by the person. We sidestep these matters here and treat agency in a simpler way that we hope does not misrepresent its fundamental qualities.

Table 14.1 Five Gateways to Successful Cognition

<i>Gateway</i>	<i>Description</i>
Register conditions	Does the learner identify conditions (IFs) and accurately instantiate their values?
Valid interpretation	Does the profile of conditions match an appropriate action?
Capability	Can the learner execute the action (THEN)?
Motivation	Will the learner spend resources to execute the action?
Capacity + opportunity	Is there cognitive capacity and environmental opportunity to execute the action?

Phases of Self-Regulated Learning

Winne and Hadwin (1998; see also Winne, 2001) described SRL as unfolding over four flexibly sequenced phases of recursive cognition (see Figure 14.1). By recursive, they meant that the results of events in any phase can feed into metacognitive monitoring and metacognitive control in any other phase.

Their model is strongly entwined with the assumption that learners are agents. Agents observe and intervene in their environments by setting goals and acting to reach them. By metacognitively monitoring qualities of how plans unfold in action, e.g., rates of progress and introspective judgments of accomplishments, learners may identify discrepancies between goals and achievements or between plans and executed processes. They then may exercise metacognitive control by revising goals, adapting plans or changing operations. In addition to attending to the task at hand, more effective self-regulated learners also

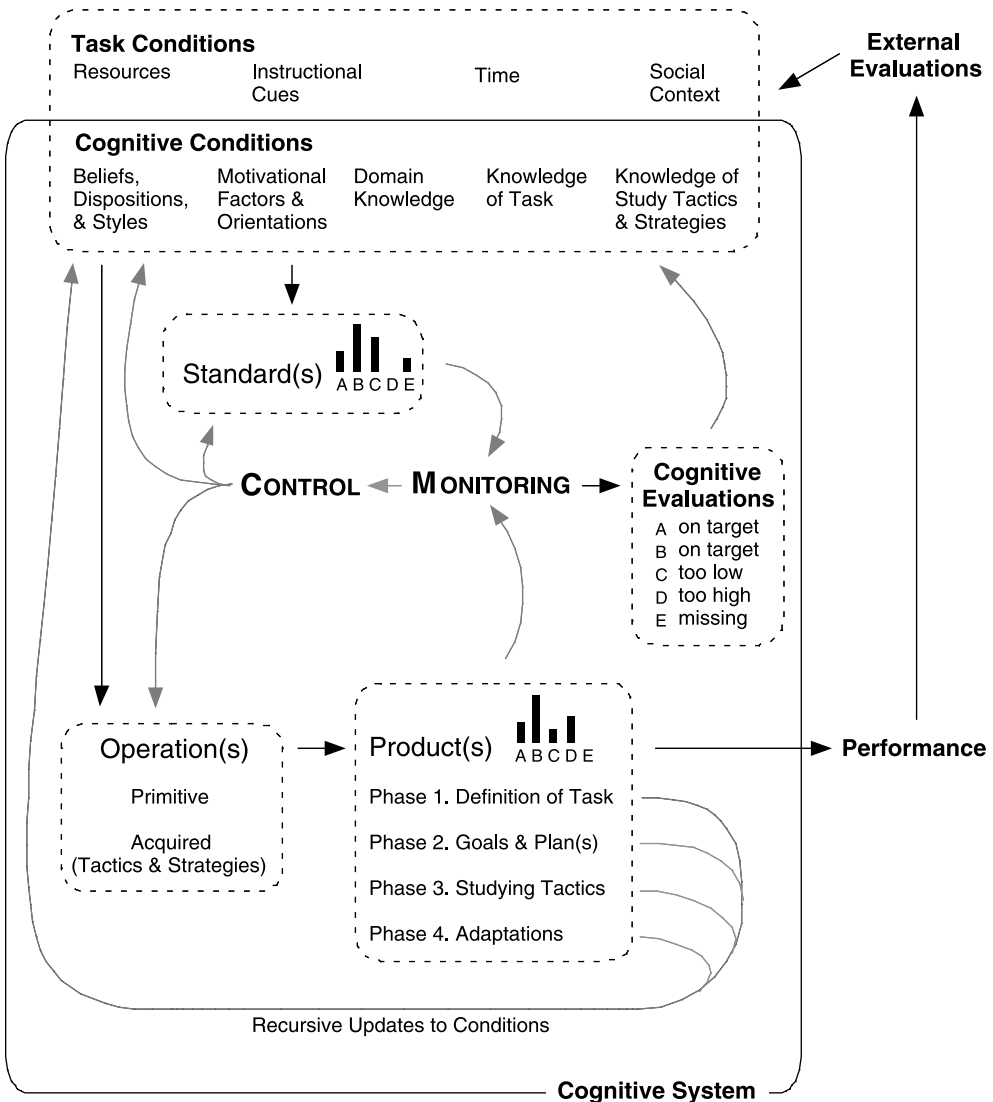


Figure 14.1 Winne and Hadwin’s (1998) four-phase model of self-regulated learning.

develop schemas and scripts for future tasks that are grounded in experience. In this sense, self-regulating learners strategically experiment with learning, fine tuning it over time, and across tasks and contexts (Winne, 1997, 2006).

In the first phase of Winne and Hadwin's model, learners scan their environment for information to construct an idiosyncratic profile of a task. This profile intrinsically blends "objective" perceptions with affective perceptions about the task (Pintrich, 2003) and other motivational information, such as judgments about self-efficacy. The result of this phase is an idiosyncratic description of the task as the learner sees it. After framing the task, learners set goals and plan work on the task relative to the framework established in phase one. Learners consider what they were assigned to achieve and weigh what they want to achieve (Pintrich, 2003). As a result, they set one or several goals. In the enacting phase, learners activate tactics and strategies to take steps toward goals. Because learners often hold multiple goals, successfully engaging in a task can entail choosing and merging methods carefully. In phase four, learners deliberate about whether and how to change their methods for completing tasks.

The adaptation in phase four can be a quite wide-ranging probe into any factor a learner can control about a task. This includes: conditions, both internal (cognitive, motivational) and external; operations, specifically which are chosen or whether typical operations should be upgraded; standards used in metacognitive monitoring; and features of evaluations, such as the latitude allowed for meeting standards or the frequency and schedule for evaluating operations and products. We adopt Winne's (2001) acronym COPES—conditions, operations, products, evaluations, standards—to synthesize these topics in one handy unit.

Challenges to Successful Metacognition and SRL

The model of SRL that we sketched earlier and its hub, metacognition, is an ideal. Being human, learners exhibit a multiplicity of non-ideal characteristics. Individually and collectively, these pose obstacles to successful metacognitive monitoring and metacognitive control in a time frame measured instance by instance; and they raise substantial barriers to learners' quests to self-regulate learning productively. Here, we present a brief catalog of these issues.

Learners are Poor Judges of What They Know

A considerable volume of research has investigated whether learners can accurately judge what they know—their metacomprehension—and factors that affect judgments of learning (JOL) (see, for example, Dunlosky & Lipko, 2007). A distinction can be made regarding whether learners judge their comprehension bit by bit (e.g., learning objective by learning objective)—termed resolution—or at an aggregate level, over all learning objectives—called calibration. While findings vary somewhat depending on type of material (e.g., word lists, paragraphs), learners' metacomprehension is often quite poor. When metacomprehension is measured using correlational statistics, ranges of near zero to +0.50 are common. Very often, learners' calibration is moderately to substantially optimistic—they judge they know material that they do not.

When learners fail to recognize occasions where metacognitive control might be helpful, they suffer a production deficiency: IFs are not noticed, so THENS are not enacted; or, in the case of overconfidence about what a learner knows, IFs are inaccurately instantiated, so inappropriate THENS are enacted. These are cases where learners fail to pass the first gate necessary to realize productive SRL.

Learners are Unschooled about Tools for Studying Effectively

To exercise metacognitive control, learners need to be able to apply effective study tactics and to exercise strategic judgment about the effects of patterns of study tactics. Some tactics are simple. For example, spacing study periods and delaying review after a first study session are very robust and effective simple tactics (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Thomas and McDaniel (2007, Table 1, p. 213) cite evidence for 11 other relatively basic tactics that enrich and deepen encoding: generating delayed summaries, generating diagrams, generating questions, re-reading, letter re-insertion (when letters are randomly deleted from text), examining inserted questions, providing perspective on content, elaborated guided reading, summarizing, generating keywords delayed from initial studying, and answering embedded questions. Others can also be identified.

Scores of studies demonstrate moderate to substantial gains in achievement when learners are taught to use these tactics. The literature on teaching reading comprehension strategies sheds light on why—without such instruction, learners fail to pass the first four gateways to productive metacognition (Garner & Alexander, 1989; Paris, Wasik, & Turner, 1991) because they miss cues for tactics, misinterpret cues, lack skills for studying, or are not motivated to pursue plausible and effective study tactics. Regrettably, the research in this area also documents a serious problem. The success of interventions shows that learners in control and comparison groups are substantially under-skilled about tools for studying effectively. Moreover, even when learners have experience that simple tools, such as clustering information into useful categories, are effective, they may abandon those tools on the basis of beliefs. We address this next.

Learners' Belief Systems can Impede Productive SRL

One key feature of Winne and Hadwin's model of SRL is that internal conditions—psychological constructs such as motivations, beliefs, and self-appraisals—are factors that influence metacognitive monitoring and control on which SRL pivots. For example, some learners engage in self-handicapping by creating impediments to performing successfully on tasks they consider important (Urduan & Midgley, 2001, p. 116). Choosing not to study the night before an exam provides an “out” if performance turns out to be poor—“I could have aced it, but I was busy with other stuff.” And, if by a stroke of luck performance turns out well, there is a comforting attribution, “Even though I didn't study, I aced it! Pretty smart, eh?”

Other belief systems affect SRL. For example, learners hold beliefs about what knowledge is and how knowledge can be manipulated, so-called epistemological beliefs. Learners with “sophisticated” epistemological beliefs view knowledge as elastic and uncertain. They expect learning to require effort, realize errors are opportunities to generate knowledge and seek explanations rather than just focus on facts. Learners with less sophisticated beliefs hold opposing views.

Beliefs affect how deeply learners process texts (Chan & Sachs, 2001) and, in fact, all phases of SRL. For example, Stahl, Pieschl, & Bromme (2006) studied relations among learners' epistemological beliefs, their calibration of task complexity and goal setting. They found learners with more sophisticated beliefs more successfully monitored task complexity, and set goals and developed plans for studying that better matched the tasks. Those with less sophisticated epistemological beliefs fared less well. Schommer, Crouse and Rhodes (1992) showed that undergraduates with less sophisticated epistemological beliefs reported they would use study tactics that process information relatively shallowly. They achieved less than students with more sophisticated epistemological beliefs.

Beliefs extend to views about how well study tactics work and how much effort they take to use—the cost–benefit or utility of tactics. Even when undergraduates perceive a tactic is effective, they may reject it if immediate past experience generates perceptions it is too effortful (Rabinowitz, Freeman, & Cohen, 1993).

Learners Struggle with Human Biases, Probabilistic Heuristics, and Difficult Reasoning

Learners can be taught new approaches to self-regulating learning but this is not common enough. In this light, Winne (1997) suggested a metaphor for how learners may become more productive in SRL—they embark on a self-directed program of research to uncover for themselves “what works.”

Prior knowledge and beliefs serve as a student’s initial theory, couched in a historically developed, personal paradigm about what learning is, methods for learning, and how this view of learning can be updated by new information. Goals correspond to hypotheses. Enacting tactics “runs” the experiment, generating data. Analyzing (monitoring) data relative to the paradigm and the specific hypothesis generates information that students can use to tune the paradigm and to frame designs for new experiments. The purpose of all this seems to be furthering one’s power to explain, predict, and control—to learn more, to avoid embarrassing consequences, or just to complete tasks with minimal effort. (p. 398)

Learners experience at least three important challenges in this quest. First, to develop expertise in a discipline, such as experimenting or in complex SRL, learners need an estimated 10,000 hours of deliberate practice:

. . . a highly structured activity, the explicit goal of which is to improve performance. Specific tasks are invented to overcome weakness, and performance is carefully monitored to provide cues for ways to improve it . . . deliberate practice requires effort and is not enjoyable. Individuals are motivated to practice because practice improves performance. (Ericsson, Krampe, & Tesch-Römer, 1993, p. 368)

But learners don’t spend nearly enough time on homework and in-class activities to meet this threshold (see Winne, 1997, p. 401).

Second, young learners have difficulty remembering steps they take in simple procedures. Without accurate records of what they have done, even when they accurately judge what works and what doesn’t, their research into metacognitive monitoring and control is not efficiently cumulative. This is a challenge with which even older learners struggle, for example, in learning procedures by studying worked examples (McLaren, Lim, Gagnon, Yaron, & Koedinger, 2006).

Third, learners of all ages have significant difficulty engaging in accurate reasoning. They use a handful of heuristics that are not always accurate (see Baron, 2000, Chapter 6). For example, judgments about events often are based on an easily recalled example, but that example is often not representative (the availability heuristic). Of great importance for reasoning about production systems, learners are not proficient at judgments that involve conditional probabilities (the representativeness heuristic), that is, under which particular conditions (IFs) does an action (THEN) “work.” In general, learners generate illusory correlations because they weigh data more heavily when it

matches their beliefs than when it challenges them. The result is a biased theory about what works.

Learners May Not Be Able to Benefit from Errors

Errors are common when learning a complex skill. Errors can lead to significant self-regulatory difficulties when the learner does not know how to correct them or does not understand why they occurred. Frequent failure of this sort decreases self-efficacy. This is important because social cognitive theory and a host of research studies point with stark certainty to the harmful motivational and performance effects of low personal efficacy (Bandura & Locke, 2003). The upshot is that, when learners perceive an inability to learn from or to recover from errors, they experience increased stress and anxiety, set less ambitious goals, and quit sooner.

Psychologists have successfully developed designs for errorless learning that avoid the implicitly punishing consequences of failure (Mueller, Palkovic, & Maynard, 2007). But while errorless learning may be an option for acquiring simple skills or parts of complex skills, errors and their resulting effects may be inevitable in complex learning environments and in the real world to which most formally acquired concepts and skills must transfer. Students must usually learn to recognize, correct, and recover from errors when they occur: The medical practitioner who diagnoses a patient must stay alert to the possibility that a diagnostic error was made so that she can recognize and correct the misdiagnosis as the patient responds to treatment. Recent research indicates a correlation between a mastery learning goal orientation and a willingness to risk errors (Arenas, Tabernero, & Briones, 2006). Students with a low mastery learning orientation may “play it safe” in complex learning environments and avoid experiencing the error conditions that they may later encounter in real life. In sum, errors are important opportunities for learning yet they may be maladaptively avoided by some learners and lead to damaging motivational effects in other learners.

Learners Don't Often Seek Help Usefully

When learners sense their learning is not going well, perhaps they should just ask for help. This simple prescription can fail in multiple ways. First, as we reviewed earlier with respect to learners' metacomprehension, learners don't always realize they need help. Second, they may actively conceal academic difficulties and avoid available help (Ryan, Patrick, & Shim, 2005). Finally, when learners do use help, the research shows they often use it ineffectively (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003).

Rather than an isolated component of self-regulation, the tendency to solicit help when needed is critically related to other aspects of self-regulation and academic success. In elementary students, help-avoidance is associated with motivationally debilitating performance avoidance goals (Ryan, Patrick, & Shim, 2005) while help-seeking predicts engagement with learning (Marchand & Skinner, 2007). Despite its importance as a learning strategy, help-seeking decreases during adolescence (Wigfield, Eccles, Schiefele, Roeser, & Davis-Kean, 2006).

Can Computing Technologies Help?

Winne (1989; see also Winne & Marx, 1982) posited that instructional effectiveness is a goal mutually sought by two intelligences. The instructor intelligently embeds cues in instructional presentations and activities to guide learners' cognition and metacognition.

Each learner intelligently strives to make sense of the instructor's communications at two levels: the object level of what is to be learned and the operational level, the activities that generate learning. By this analysis, software technologies theoretically might help learners engage in more productive SRL by providing cognitive tools and other affordances that help learners identify what is to be learned and how to learn it. Two critical factors that determine the potential success of this proposal are (a) the data that software can gather and (b) methods by which the software can analyze data and report results of those analyses to learners as feedback.

What Data Can Software Gather?

Software that logs learners' interactions as they use the software's features in learning activities has become common. Three fundamentally distinct kinds of information can be logged. The first kind of log data is records of which features of the software the learner uses. Features can range from basic ones, such as clicking a close box that puts away a window; to complex ones, such as engaging with a wizard to forge a complicated Boolean design for searching information within conditionally identified files or directories. Some features are presented semantically, such as actions described by menu options or choices expressed by clicking a button with a label that answers a question posed in a dialog box. Log data about learners' uses of these features can enrich the information value of data by virtue of the semantics inherent in the feature. What all these data have in common is that they identify how the learner is operating when using the software. These data describe processes.

The second kind of information that can be logged as data is the information on which the learner operates. There are two subtypes here: (a) information to be learned—content; and (b) information intended to guide the learner about how to learn—instructional cues. For example, suppose a learner uses a system's tools to copy text from a source, paste it into a field, then type text in the field to elaborate that quotation. The learner has created an annotation. The text selected and the text generated by the learner are two instances of data.

The third kind of information logged by modern instructional software systems is temporal data. These identify when an event happens. By stamping each event along a time line, it is possible to build up representations that identify durations of single events, such as deciding; and spans between events, such as re-studying previously viewed content (Winne, Jamieson-Noel, & Muis, 2002; Winne & Perry, 2000).

By combining all three types of data, very complex and variably interpretable characterizations can be developed of how learners go about their work during learning.

What Can Software Do Better than a Learner?

As we reviewed earlier, students face many psychological challenges in becoming productive, self-regulated learners. Why might software help them overcome these challenges?

First, learners draw biased samples of experience, in part because they cannot observe every event and in part because human memory is imperfect. Learners' perceptions are influenced by a host of factors, including the pace of events, preoccupation with some matters over others, and bias. With respect to the data that software can log, these shortcomings are erased. Of course, the critical question is whether the data that software can log are sufficient for generating accurate models of events.

Second, learners appear not to be sufficiently reliable "calculators" when they examine data in mind or at hand. While often consistent in their approaches to processing data,

solutions they generate are often biased in directions that are contrary to optimizing learning. Moreover, learners are very challenged to carry out analyses. They have trouble computing frequency counts, correlations between variables, and critically for self-regulation, conditional probabilities that link IFs and THENs. Perhaps the most appropriate computation for SRL, Bayesian network analyses, are impossible without support. In contrast, computations such as these are software's strong suit.

Third, collecting some kinds of data that may reveal significant features of learning processes can interfere with learners' focus on carrying out those processes. That is, learning and simultaneously collecting data about learning presents learners with a quite difficult dual task. Software technologies, however, can easily manage this dual tasking at the human pace of learning.

Fourth, learners are rather unreliable observers. Software technologies are, for all intents and purposes, perfectly reliable. The challenge, however, is to gather data not only because it is reliable but because it can reveal or contribute to forecasting important features of learning processes.

Fifth, learners forget and misconstrue what they do. Logs of learner engagements, however, are complete and accurate. Thus, when learners succeed or experience a particular kind of satisfaction over a series of learning events, the data logged by software can literally recreate the path the learner followed.

How Can Software Facilitate Metacomprehension?

To pass the first gateway, registering features in the learning environment, learners must accurately self-evaluate their knowledge and understanding. Learners' accuracy in assessing what they know increases when they try to use that knowledge to complete a task. For instance, learners make better judgments of how much they learned from a text when, with some delay after reading, they try to generate keywords characterizing the text (Thiede, Dunlosky, Griffin, & Wiley, 2005). Tasks that improve judgment of learning, such as spaced practice or generating summaries, give learners both a sample of performance and a standard against which to judge that performance. Significantly, "metacomprehension prediction accuracy will improve to the extent that people engage in enriched-encoding activities" (Thomas & McDaniel, 2007, p. 212), which are the same types of deep learning activities that are known to promote learning itself.

Fundamentally, software can facilitate metacomprehension in two ways: first, by assisting the learner to engage in learning activities in which they deeply process information; second, by prompting them to self-assess metacomprehension as well as the knowledge gained (comprehension) and practiced through those activities. Although they hold no monopoly on prompting deep learning activities, software technologies do uniquely offer certain features such as threaded discussions, in which students argue over the main points in a reading; and structured instant messaging, in which students adopt identified roles and use pre-designed message starters (Hadwin, Oshige, Gress, & Winne, in press). Of course, software support for meaningful learning is not limited to collaborative activities, as demonstrated by the many examples of educational simulations, interactive learning environments, concept mapping tools, and so on.

Software can facilitate learners' metacognitive monitoring by providing opportunities to record regularly and conveniently their perceived ability or performance level, perhaps in the form of graphic "knowledge meters" that feature adjustable dials or sliders. Multiple learning goals or dimensions of knowledge could be represented by multiple meters. Such reminders can increase chances that learners attend to opportunities for metacognitive monitoring and perceive dimensions (standards) to use in monitoring. The same software

could generate graphs that show how judgments of ability change over time, presumably rising with progress toward learning goals. This enhances learners' capabilities to track what they do.

The concept of accurate or calibrated metacomprehension implies a standard such as a personally constructed performance goal; or, for many students in a formal educational system, scores on tests, assignments and other assessments. Because computer-based assessment systems collect and report performance scores, and could also collect students' predictions of their performance (judgments of learning), these systems and other software have the potential to offer metacognitive feedback that students could use to increase the accuracy of their self-assessments. Psychometric technologies and methods are continuing to develop, increasing the likelihood that computerized testing and assessment will be gradually introduced into the academic lives of students everywhere. Adding to the well-established models of adaptive testing, advances are being made in delivering tests over the internet (Bartram & Hambleton, 2006) and automated scoring of complex tasks (Williamson, Mislevy, & Bejar, 2006). What feedback might computer-based testing systems give, beyond performance scores, that could raise students' metacognitive accuracy? Such systems might regularly ask students to judge their performance before the test, and then give feedback on the accuracy of the prediction in the form of a difference metric. These events address the first three gateways to more productive metacognitive engagement: registering conditions, validly interpreting conditions, and capability to carry out action.

Other types of educational software, beyond those dedicated to assessment, gather data on learner performance and use them to construct models of learners' knowledge and ability. For example, intelligent tutoring systems often include a learner model as one structural sub-component. To increase metacognitive accuracy, such systems may be able to make the model transparently available to learners. Although learner models may consist of production rules or data tables that are not readable by learners, it may be possible to translate the models into meaningful reports that boost their metacognitive awareness and help them overcome limitations in analyzing such data.

How Can Software Teach Learning Tactics and Strategies?

Another requirement for passing the first gateway, registering conditions, is being aware of the actions one has performed and is performing. This entails accurate self-monitoring. Self-monitoring is a key component in cognitive behavior therapy, and is broadly used by programs in which patients self-manage behavioral change (Farmer & Chapman, 2008). It has seen significant usage in cognitive therapies for addressing students' problems with academic skills, in which students are taught to record and assess aspects of their academic performance (Bradley-Klug & Shapiro, 2003; Shapiro, Durnan, & Post, 2002). In a typical application of self-monitoring, elementary students with ADHD might be taught to count and graph the number of times they used a spelling practice procedure (Harris, Friedlander, Saddler, Frizzelle, & Graham, 2005). Various types of self-monitoring enabled by the record-keeping capabilities of computers may be effective in helping students to improve their learning strategies.

We have developed software called gStudy that supports self-regulated learning (Nesbit & Winne, 2007; Perry & Winne, 2006; Winne et al., 2006). The cognitive tools provided by gStudy allow learners to operate on information in web pages using a variety of tactics such as tagging, making notes in structured forms, and visually mapping concepts. Our experience using gStudy in university classes is that merely showing students how to use the cognitive tools does not lead them to adopt the tactics or change

students' learning strategies—they fail to pass the fourth gateway of motivation. We have found that through specially designed course activities, instructors can help students toward more frequent use of the tools and tactics in their coursework (Nesbit et al., 2006). In addition to efforts by instructors, what can software do to contribute to such changes?

Software such as gStudy has the potential to show learners an aggregate view of their past learning tactics in the form of graphs, tables, and activity timelines. It can show them how many tags, summaries, comments, questions, and so on they created while working with multimedia documents; how much time they spent working with different sections of content; and how the types of activities changed over time. But the mere availability of such information does not constitute self-monitoring. We must also build in incentives and cues for learners to attend to such reports and use them as input to SRL.

Records of individual students' use of cognitive tools during an assigned activity could be made available to the teacher, who might include the use of particular learning tactics as part of the assignment. As an extrinsic reward, a portion of the grade for the assignment could be based on the students' recorded use of learning tactics. Teachers could also use such records as the basis for individual discussion with students about their learning strategies. Also aggregated, anonymous records for a whole class could be made available to individual students so they can compare their learning strategies to those of peers. Both sharing with teacher and peer comparison bring into play potentially powerful sociogenic motivations.

To pass the second gateway, validly interpreting conditions, learners must know and remain aware of which actions are appropriate for present conditions. Sometimes learners don't know where to start, or they miss opportunities to use cognitive tools. An automated pedagogical tutor with access to the learner's record of learning tactic usage could prompt the students with choices. For example, the tutor might advise learners that they have been doing mostly highlighting in a document they are studying and should consider making some notes.

Schwonke, Hauser, Nückles and Renkl (2006) aimed to foster learners' writing of a learning protocol which is "a written explication of one's own learning processes and outcome. When this occurs over an extended period of time, it is called a 'learning diary' " (p. 78). They provided university undergraduates access to eHELP, a computer-based learning environment that adapted prompts for writing learning protocols based on data gathered in self-reports. Some students answered a 15-item questionnaire describing their typical usage of organizational, elaborative and metacognitive learning strategies. Others took a test of meta-knowledge where they read nine scenarios describing learning situations and rated the appropriateness of five given strategies. The prompts that eHELP provided subsequently, when students wrote learning protocols, varied according to the pre-measure students wrote. For those writing the questionnaire, prompts corresponded to the five lowest-rated learning strategies in the questionnaire. For students who answered the meta-knowledge test, prompts were given when students assigned an inappropriate strategy to a scenario. In other words, the prompts brought to students' attention various operations (strategies) under particular conditions of "need," thereby modeling metacognitive monitoring and suggesting choices of metacognitive control. In comparison to students who received no prompts or random prompts, both questionnaire-based and scenario-based prompts led to greater indications of metacognition in learning protocols. And, among students who received prompts, or prompts adapted on the basis of low self-reported usage or low judgments of appropriateness, adapted prompts were judged more supportive.

How Can Software Help Learners Benefit from Errors?

Researchers have examined the effects of instructional methods that (a) allow learners to make and correct errors, (b) ensure that learners generate a sufficient number of errors, and (c) guide learners to make and correct common errors. Compared with methods that avoid learner errors, methods providing for learning from errors in the context of complex skills help learners recognize when they have made a mistake, develop corrective strategies, and more successfully transfer acquired skills (Dormann & Frese, 1994; Gully, Payne, Kiechel Koles, & Whiteman, 2002; Lorenzet, Salas, & Tannenbaum, 2005). Such methods may be especially important for learners with low mastery learning orientation who actively avoid exploration in complex learning environments. Such software, specifically, can address the first and the second gateways of noticing errors and perceiving information in errors.

Despite the demonstrated advantages of exposing learners to errors and teaching them corrective strategies, it is not always practical for teachers to provide full support for errors. That is, classroom environments fail to pass the fifth gateway of providing opportunity without overloading cognitive capacity.

Software can help by providing learning environments in which the learner can exercise significant self-direction and also receive exposure to important error conditions with assistance needed to recover and learn from them. First, the software can keep track of the number and types of errors made by learners. Second, when it detects the learner has not been exposed to an important error condition, it may introduce a challenge that is likely to cause the student to make the error. Alternatively it may directly instruct the student to make a response that produces the error (e.g., "Let's see what happens when you lower the coolant levels in the reactor core."). Third, it can show the consequences of the error (e.g., "The reactor core is overheating."). Fourth, it can provide an explanation for why the student's response was an error. And finally, it can explain how to recover from the error. These events set the stage for passing the first through third gateways: registering conditions, validly interpreting conditions, and capability to carry out action.

Tracking, inducing, and explaining errors in complex learning environments are characteristics of a guided discovery approach to educational software design. Research by Moreno (2004) demonstrated how one of these elements, explanations of errors, can operate in guided discovery software. Moreno had learners work with multimedia software that taught them to design plants for survival under different environmental conditions. The learners selected plant features and received spoken feedback about the correctness of each choice from a pedagogical agent presented by the software. Approximately half the learners also received an explanation of the suitability of their choices (e.g., "Your deep roots will not help your plant collect the scarce rain that is on the surface."). Although learners receiving the explanation treatment spent approximately the same amount of time working with the software, they obtained higher scores on retention and transfer tests, and rated the software as more helpful. Because knowing the right answer is not the same as knowing all the reasons *why* it is the right answer, we speculate that, when using software like that studied by Moreno, learners with a high mastery learning orientation may intentionally make incorrect responses so they can learn from the explanations offered as error feedback. Sophisticated guided discovery software would ensure that even those learners with low mastery learning orientation would encounter important error conditions and their explanations.

How Can Software Foster Adaptive Help-Seeking?

Although academic help-seeking and its role in self-regulated learning have received significant scholarly attention (e.g., Karabenick, 1998), research on interventions that teach effective help-seeking strategies are still nascent. Even less consideration has been given to the functions software might serve in guiding learners toward adopting help-seeking strategies. We believe there are two types of help-seeking that such software might address: (a) help received from humans, such as peers and teachers; and (b) help received from automated systems, particularly the help features embedded in digital learning environments.

In the high-bandwidth, wireless communication environment now prevalent in the developed world, much of the help that learners receive from peers and teachers is or has the potential to be mediated by digital technology. Whether the communication is by instant messaging, email, voice, or video, technology's position as mediator of these communications is prime territory for facilitating the exchange of help in a variety of ways. Software could identify potential helpers, assist the learner to select a helper, and make it easier to frame the request for help by packaging the contextual information that the helper will need to respond effectively. An example of a system that fulfills some of these functions is I-Help (Bull, Greer, & McCalla, 2003; Greer et al., 2001), networked software that brokers academic help from peers in an online environment. I-Help seeks to identify helpers that match the learner's needs and preferences on several psychosocial and logistical variables. Theoretically, I-Help should be able to radically reduce the difficulty and effort involved in accessing help; it can also prompt learners to consider metacognitively the kinds of help from which they might profit, moving them through the second gateway, validly interpreting conditions, to productive metacognition.

Software that makes accessing help easier, such as I-Help, also eases the learner past our third gateway, capability to carry out action; and fifth gateway, having opportunity without overloading cognitive capacity. It seems less certain, however, that such systems would increase the learner's awareness of when help is needed (gateway 1) or increase their motivational tendency to seek help when it is needed (gateway 4). To help learners pass those gateways we need software that intervenes with recommendations and explanations and then fades these as the learner develops autonomy. If the software could model a learner's attempts to solve a problem, then it could recommend seeking help when it detects that progress has stalled and give reasons why seeking help is the right thing to do. The recommendation might be something like:

I see that you are stuck on this question. Many students experience difficulty with questions like this. You will save time and learn more efficiently if you ask for help. Emma, who is now online, likely can help you with this problem.

In addition to mediating online exchanges, software of this type may be able to facilitate teachers' provision of help in classrooms. Such a classroom help system could keep track of the attention teachers give to individual students, queue students to receive teacher help, advise teachers on students' priority for receiving help, and generally operate to balance the often inequitable dispensing of academic assistance. Simply receiving more frequent help from a teacher is likely one of the most effective ways to reduce maladaptive tendencies toward help-avoidance. Systematic improvements in the distribution of academic assistance would indirectly expose students to pragmatic standards for judging when to seek help.

We are not aware of any software that uses a learner model to actively instruct seeking help from peers or a teacher, but extending ideas first researched by Luckin and Hammerton

(2002), Alevan, McLaren, Roll and Koedinger (2006) have made remarkable progress in guiding high-school geometry learners about when and how they access help from a geometry tutoring system. Their Help Tutor used 57 production rules that implemented an optimized model of productive and unproductive help-seeking as a supplement to a cognitive tutor (see Anderson, Corbett, Koedinger, & Pelletier, 1995) in which learners are posed real-world problems in geometry and provided real-time individualized feedback plus on-demand advice about problem-solving activities. The geometry tutoring system itself has two embedded help features: (a) context-sensitive hints giving increasingly specific recommendations about how to complete the current step in a geometry problem, and (b) a context-insensitive glossary that provides definitions of geometry rules. As learners work, the Help Tutor makes recommendations to steer them toward asking for a context-sensitive hint if they have no idea what to do next, or toward the glossary if they have some sense of what to do but are not certain. To judge whether the learner should seek help and, if so, what type, the Help Tutor uses information such as the learner's past performance and the time taken to respond to a step in the problem. The Help Tutor's model of the learner considered kinds of help available in relation to conditions of learner mastery and time spent in particular learning activities, such as thinking about geometry knowledge that might be applicable at a point in solving a problem. These are steps toward productive metacognition through the first gateway of registering conditions. Through the recommendations it makes to learners, the Help Tutor mitigates maladaptive strategies such as help avoidance (e.g., guessing rather than consulting help) or "help abuse," which is repeatedly accessing the context-sensitive help feature until it practically gives away the solution.

Software capable of teaching effective strategies for seeking help, whether from humans or digital systems, would need to include: (1) a model of the learner's abilities in the subject domain, so the degree and type of help needed can be accurately judged; (2) a model of appropriate help-seeking strategies; (3) a model of the actual help-seeking practices of the individual learner; (4) a pedagogical model that heuristically regulates how often and under which conditions recommendations about help-seeking should be made to the learner. The latter model is important to prevent learners from rejecting or ignoring the recommendations because they occur too frequently or add to the cognitive load already incurred by the main task. Although the same structural framework might be used for teaching learners to seek help from humans and computers, the learning goals would differ substantially. An important goal for teaching learners to seek help from teachers and peers is moving them through gateway four by overcoming the debilitating sociogenic motivations related to concealment of inability, performance avoidance, and fear of failure.

How can Software Motivate Learners to Self-Regulate?

To pass the fourth gateway of motivation learners must be sufficiently motivated to perform self-regulatory actions they deem to be applicable. It may appear that at this gateway we have finally met too great a challenge. How can technology enhance motivation, a psychological factor that is often seen as an inherent personal disposition or a dynamic construct emerging from the social environment? While acknowledging there may be as many approaches to this question as there are motivational theories, we choose to explore how software can contribute by considering the role of self-efficacy for self-regulation, which is one's perception of one's ability to plan, monitor, and regulate while completing a task (Bandura, 1997). Self-efficacy for self-regulation can be as strong a predictor of achievement as direct self-efficacy for achievement (e.g., Mill, Pajares, & Herron, 2007).

One of the ways that teachers can increase learner self-efficacy is through verbal persuasion, which is usually encouragement to take small risks and invest effort. Software can amplify teachers' efforts in this regard by facilitating and increasing effective teacher-learner communication. There is evidence that electronically delivered verbal persuasion can boost learner self-efficacy for achievement. Jackson (2002) found that email messages from an instructor designed to boost undergraduate students' self-efficacy raised both self-reports of self-efficacy and subsequent exam performance. Because the students had already completed three graded activities, in constructing the email sent to each student, the instructor was able to emphasize the learner's relatively successful past performances. Of course, appropriately designed software could automate the task of assembling and analyzing performance data needed to construct such messages and in so doing it could lower the costs incurred by the instructor in increasing verbal persuasion.

Fostering self-efficacy for self-regulation is more difficult because it pivots on knowledge of learners' past success in monitoring and regulating their learning, data which are not usually available to teachers. We have already seen, though, how software systems can collect, analyze, and report such data. If learners have previously been asked to judge their level of understanding before taking a test, then a report can be compiled for each learner showing the judgments that were most accurate. This may be a first step to passing through gate number three, improving learners' capabilities to accurately metacognitively monitor. If learners' study tactics have been recorded, then a report can identify the study sessions or content sections in which the learner most assiduously applied certain tactics. Software can automatically present analyses to teachers that they are unlikely to conduct by hand and can present graphs and tables that identify trends or relationships that emphasize students' successes and failures in self-regulated learning. That is, these reports might be able to characterize conditions where students miss opportunities to apply productive self-regulation at gate number one. Teachers can then use this information to motivate students through verbal persuasion. We anticipate a day when a teacher can send an audio message to a student that says something like the following:

Hi Julian. I notice that you have been reading Section 4 on the Industrial Revolution to prepare for the class debate. Now would be a good time to figure out which concepts and connections you do not completely understand in that section and write notes that explain them. You already showed you could do these things last week when you accurately estimated your knowledge of the Napoleonic Wars and in the section on the French Revolution where you made notes that mentioned almost every major concept and event. Anyway, I hope you enjoy this section.

Another way that software can foster self-efficacy for self-regulation is providing feedback to the learner on associations between past self-regulatory actions and their outcomes. Because the effects of self-regulation are delayed and uncertain, it is easy for learners to sometimes conclude that their efforts to monitor learning and act strategically are to no avail. Thus, they can misperceive information needed to pass through gateway two, validly interpreting conditions. Learners trying to develop SRL skills who make great efforts to test themselves and review accordingly in just one part of the assigned material may do poorly on a test that covers all the material. In this case the crude feedback usually received—a graded response sheet—may considerably dampen any burgeoning self-efficacy for SRL. But if the learner instead receives a report that shows a link between strategic action in a few parts of the material and success on items testing those parts, the motivational impact of the feedback may turn from negative to positive. Software is not only capable of finding such associations but also of graphically depicting

them to the learner in a display that is more clear and convincing than any descriptive text or numeric table.

These are two examples of possible ways that software might ease a learner's path across the fourth gateway of motivation. In general, software can support motivation to self-regulate by accurately logging, analyzing, and presenting information in the service of any interventions that are known to motivate learners. Because motivational interventions will not occur if the cost of processing the required information is too high, appropriately designed software could radically expand the motivational support learners receive for metacognitive action.

Toward Embedded and Pervasive Metacognitive Tools

It is probably not practical to develop independent software systems dedicated solely to supporting and tutoring metacognition and SRL because independent systems would not have access to data about learner performance and learners would have little incentive to use software that did not directly provide cognitive support in assigned tasks. Instead, software tools that develop learners' metacognition should piggyback on other educational software as symbiotic modules, using the data and learner models generated by these purpose-built forms of educational software.

We have discussed several examples of how this could work: Metacognitive modules allowing learners to predict scores could piggyback on testing systems; help tutors could piggyback on tutoring systems; error guidance modules could piggyback on interactive learning environments and educational simulations; tactical recommendation agents could piggyback on educational annotation and concept visualization systems; and aids to learners' episodic recollection of learning events generated from logs could piggyback on all types of educational software.

Of course, this list is not exhaustive. It seems plausible to us that every possible type of educational software could be extended to promote learners' metacognitive development. We urge every designer of educational software to consider ways that learners using their software can be helped toward greater awareness of the learning process and their actions within that process. Harking back to the comments by the Institute's panel (Pashler et al., 2007) quoted in our introduction, if self-regulation of learning is an important ability with lifelong effects, then the features software developers add to develop that ability may be their greatest contribution.

References

- Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help-seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16*, 101–130.
- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research, 73*, 277–320.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*, 167–207.
- Arenas, A., Taberero, C., & Briones, E. (2006). Effects of goal orientation, error orientation, and self-efficacy on performance in an uncertain situation. *Social Behavior and Personality, 34*, 569–586.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*, 87–99.
- Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge, UK: Cambridge University Press.

- Bartram, D. & Hambleton, R. (Eds.) (2006). *Computer-based testing and the internet: Issues and advances*. Chichester, UK: Wiley.
- Bradley-Klug, K. L., & Shapiro, E. S. (2003). Treatment of academic skills problems. In M. A. Reinecke, F. M. Dattilio, & A. Freeman (Eds.), *Cognitive therapy with children and adolescents: A casebook for clinical practice* (pp. 281–303) (2nd ed.). New York: Guilford.
- Bull, S., Greer, J., & McCalla, G. (2003). The caring personal agent. *International Journal of Artificial Intelligence in Education*, 13, 21–34.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132, 354–380.
- Chan, C. K. & Sachs, J. (2001). Beliefs about learning in childrens understanding of science texts. *Contemporary Educational Psychology*, 26, 192–210.
- Dormann, T. & Frese, M. (1994). Error training: Replication and the function of exploratory behavior. *International Journal of Human Computer Interaction*, 6, 365–372.
- Dunlosky, J. & Lipko, A. R. (2007). Metacomprehension. A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, 16, 228–232.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.
- Farmer, R. F. & Chapman, A. L. (2008). *Behavioral interventions in cognitive behavior therapy: Practical guidance for putting theory into action*. Washington, DC, US: American Psychological Association.
- Garner, R. & Alexander, P. (1989). Metacognition: Answered and unanswered questions. *Educational Psychologist*, 24, 143–158.
- Greer, J., McCalla, G., Vassileva, J., Deters, R., Bull, S., & Kettel, L. (2001). Lessons learned in deploying a multi-agent learning support system: The I-Help experience. In J. D. Moore, C. L. Redfield, & W. L. Johnson (Eds.), *Artificial intelligence in education* (pp. 410–421). IOS Press, Amsterdam.
- Gully, S. M., Payne, S. C., Kiechel Koles, K. L., & Whiteman, J. K. (2002). The impact of error training and individual differences on training outcomes: An attribute-treatment interaction perspective. *Journal of Applied Psychology*, 87, 143–155.
- Hadwin, A. F., Oshige, M., Gress, C. L. Z., & Winne, P. H. (in press). Innovative ways for using gStudy to orchestrate and research social aspects of self-regulated learning. *Computers in Human Behavior*.
- Harris, K. R., Friedlander, B. D., Saddler, B., Frizzelle, R., & Graham, S. (2005). Self-monitoring of attention versus self-monitoring of academic performance: Effects among students with ADHD in the general education classroom. *The Journal of Special Education*, 39, 145–156.
- Jackson, J. W. (2002). Enhancing self-efficacy and learning performance. *Journal of Experimental Education*, 70, 243–254.
- Jamieson-Noel, D. L. & Winne, P. H. (2003). Comparing self-reports to traces of studying behavior as representations of students' studying and achievement. *German Journal of Educational Psychology*, 17, 159–171. [*Zeitschrift für Pädagogische Psychologie*]
- Karabenick, S. A. (Ed.) (1998). *Strategic help seeking: Implications for learning and teaching*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lorenzet, S. J., Salas, E., & Tannenbaum, S. I. (2005). Benefiting from mistakes: The impact of guided errors on learning, performance, and self-efficacy. *Human Resource Development Quarterly*, 16, 301–322.
- Luckin, R. & Hammerton, L. (2002). Getting to know me: Helping learners understand their own learning needs through meta-cognitive scaffolding. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *Proceedings of Sixth International Conference on Intelligent Tutoring Systems, ITS 2002* (pp. 759–771). Berlin: Springer.
- Marchand, G. & Skinner, E. A. (2007). Motivational dynamics of children's academic help-seeking and concealment. *Journal of Educational Psychology*, 99, 65–82.
- Martin, J. (2004). Self-regulated learning, social cognitive theory, and agency. *Educational Psychologist*, 39, 135–145.
- McLaren, B. M., Lim, S., Gagnon, F., Yaron, D., & Koedinger, K. R. (2006). Studying the effects of

- personalized language and worked examples in the context of a web-based intelligent tutor. In M. Ikeda, K. Ashley, & T. Chan (Eds.), *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (pp. 318–328). New York: Springer.
- Mills, N., Pajares, F., & Herron, C. (2007). Self-efficacy of college intermediate French students: Relation to achievement and motivation. *Language Learning, 57*, 417–442.
- Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*, 99–113.
- Mueller, M. M., Palkovic, C. M., & Maynard, C. S. (2007). Errorless learning: Review and practical application for teaching children with pervasive developmental disorders. *Psychology in the Schools, 44*, 691–700.
- Nesbit, J. C. & Winne, P. H. (2007). Tools for learning in an information society. In T. Willoughby & E. Wood (Eds.), *Children's learning in a digital world* (pp. 173–195). Blackwell Publishing, Oxford, UK.
- Nesbit, J. C., Winne, P. H., Jamieson-Noel, D., Code, J., Zhou, M., MacAllister, K., Bratt, S., Wang, W., & Hadwin, A. F. (2006). Using cognitive tools in gStudy to investigate how study activities covary with achievement goals. *Journal of Educational Computing Research, 35*, 339–358.
- Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research, Vol. II* (pp. 609–640). New York: Longman.
- Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (NCER 2007–2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ncer.ed.gov>.
- Perry, N. E. & Winne, P. H. (2006). Learning from learning kits: gStudy traces of students' self-regulated engagements using software. *Educational Psychology Review, 18*, 211–228.
- Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*, 667–686.
- Rabinowitz, M., Freeman, K., & Cohen, S. (1993). Use and maintenance of strategies: The influence of accessibility to knowledge. *Journal of Educational Psychology, 84*, 211–218.
- Ryan, A. M., Patrick, H., & Shim, S.-O. (2005). Differential profiles of students identified by their teacher as having avoidant, appropriate, or dependent help-seeking tendencies in the classroom. *Journal of Educational Psychology, 97*, 275–285.
- Schommer, M., Crouse, A., & Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it is simple does not make it so. *Journal of Educational Psychology, 84*, 435–443.
- Schwonke, R., Hauser, S., Nückles, M., & Renkl, A. (2006). Enhancing computer-supported writing of learning protocols by adaptive prompts. *Computers in Human Behavior, 22*, 77–92.
- Shapiro, E. S., Durnan, S. L., & Post, E. E. (2002). Self-monitoring procedures for children and adolescents. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 433–454). Washington, DC, US: National Association of School Psychologists.
- Stahl, E., Pieschl, S., & Bromme, R. (2006). Task complexity, epistemological beliefs and metacognitive calibration: An exploratory study. *Journal of Educational Computing Research, 35*, 319–338.
- Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*, 1267–1280.
- Thomas, A. K. & McDaniel, M. A. (2007). Metacomprehension for educationally relevant materials: Dramatic effects of encoding-retrieval interactions. *Psychonomic Bulletin & Review, 14*, 212–218.
- Urdan, T. & Midgley, C. (2001). Academic self-handicapping: What we know, what more there is to learn. *Educational Psychology Review, 13*, 115.
- Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R., & Davis-Kean, P. (2006). Development of achievement motivation. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child*

psychology: Vol. 3. Social, emotional, and personality development (6th ed., pp. 933–1002). New York: Wiley.

- Williamson, D. M., Mislevy, R. J., & Bejar, I. I. (2006). *Automated scoring of complex tasks in computer-based testing*. Mahwah, NJ: Erlbaum.
- Winne, P. H. (1982). Minimizing the black box problem to enhance the validity of theories about instructional effects. *Instructional Science*, *11*, 13–28.
- Winne, P. H. (1989). Theories of instruction and of intelligence for designing artificially intelligent tutoring systems. *Educational Psychologist*, *24*, 229–259.
- Winne, P. H. (1992). State-of-the-art instructional computing systems that afford instruction and bootstrap research. In M. Jones & P. H. Winne (Eds.), *Adaptive learning environments: Foundations and frontiers* (pp. 349–380). Berlin: Springer-Verlag.
- Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology*, *89*, 397–410.
- Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 153–189). Mahwah, NJ: Lawrence Erlbaum Associates.
- Winne, P. H. (2006). How software technologies can improve research on learning and bolster school reform. *Educational Psychologist*, *41*, 5–17.
- Winne, P. H. & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence Erlbaum Associates.
- Winne, P. H. & Jamieson-Noel, D. L. (2002). Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*, 551–572.
- Winne, P. H. & Jamieson-Noel, D. L. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology*, *28*, 259–276.
- Winne, P. H., Jamieson-Noel, D. L., & Muis, K. (2002). Methodological issues and advances in researching tactics, strategies, and self-regulated learning. In P. R. Pintrich & M. L. Maehr (Eds.), *Advances in motivation and achievement: New directions in measures and methods* (Vol. 12, pp. 121–155). Greenwich, CT: JAI Press.
- Winne, P. H. & Marx, R. W. (1982). Students' and teachers' views of thinking processes for classroom learning. *Elementary School Journal*, *82*, 493–518.
- Winne, P. H., Nesbit, J. C., Kumar, V., Hadwin, A. F., Lajoie, S. P., Azevedo, R. A., & Perry, N. E. (2006). Supporting self-regulated learning with gStudy software: The Learning Kit Project. *Technology, Instruction, Cognition and Learning*, *3*(1), 105–113.
- Winne, P. H. & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). Orlando, FL: Academic Press.

15 Effective Implementation of Metacognition

Michael J. Serra and Janet Metcalfe
Columbia University

Metacognition can be described as “a critical analysis of thought,” “knowledge and cognition about cognitive phenomena” (Flavell, 1979), or simply “thinking about thinking.” It can focus on any aspect of cognition, even metacognition itself (Dunlosky, Serra, Matvey, & Rawson, 2005; Nelson & Narens, 1994). Metacognition has typically been conceptualized as involving one or more of the following aspects of a cognitive process: knowledge about that process, the monitoring of that process, and the control of that process. When optimized, these aspects of metacognition can augment performance of the target cognition, including students’ learning (Azevedo & Cromley, 2004; Winne, 1995). For this reason, many researchers have sought to train students to engage in metacognitive thinking to improve their learning (e.g., White & Frederiksen, 1998).

Unfortunately, metacognition is prone to errors and often exhibits poor accuracy, which can have a negative impact on its implementation in studying and learning (e.g., Nelson & Dunlosky, 1991; Maki, 1998b; Metcalfe & Finn, 2008a; Thiede, 1999; Thiede, Anderson, & Therriault, 2003). Although we believe that metacognition should be taught in the classroom as an important learning skill, students and teachers must be made aware of the errors that can accompany metacognition and learn ways to sidestep them. Towards this end, this chapter discusses the effective implementation of metacognition. We present an overview of metacognition and discuss how its three main components—knowledge, monitoring, and control—play a role in studying for retention and comprehension. For each component, we review some of the literature relevant to optimizing metacognition. In particular, we focus on errors/biases in monitoring, and the negative effects that poor monitoring accuracy can have on control processes. We also review some of the heuristics that people use to make their metacognitive judgments, and methods that have been shown to improve the accuracy of metacognitive judgments. We also suggest some ways to best incorporate accurate monitoring into study.

Knowledge

Metacognitive knowledge is information that one consults when thinking about a particular cognition. This knowledge can include information about the cognitive task at hand, about one’s ability to perform that task, or about potential strategies one might use to perform that task. Metacognitive knowledge informs monitoring and control, and the accuracy and efficacy of these latter two functions increase as metacognitive knowledge increases (Winne & Hadwin, 1998; see also Wilson & Brekke, 1994). Metacognitive knowledge might exert an influence in the classroom in a number of ways. For example, a student who expects an essay test instead of a multiple-choice test on a chapter might study by writing a summary of the chapter instead of attempting to answer the practice short-answer questions in the back of the chapter. He or she presumably knows that trying

to write a summary is more representative of how understanding will be tested on such a test than answering the multiple-choice questions is (knowledge affecting monitoring). Another student who knows that he or she struggles to understand mathematics lessons might plan to devote extra time to study or seek extra help (knowledge affecting control).

Students who know more about how to study and about how learning occurs (i.e., those with more metacognitive knowledge) learn better than those with less metacognitive knowledge (Winne & Hadwin, 1998). For this reason, educating students about how they learn and identifying effective (and ineffective) learning strategies for them should not only improve the accuracy of their metacognitive judgments, but should also improve their self-regulated learning (Azevedo & Cromley, 2004; White & Frederiksen, 1998). Teachers and researchers can refer to the practice guides available on the Institute of Education Sciences (IES) website for helpful reviews of the efficacy of approaches to various educational topics (e.g., encouraging girls in math and science) to become better informed themselves in terms of which instructional methods are effective and which are not. Of most relevance to the current topic will be the guide on organizing instruction and study to improve student learning (Pashler et al., 2007). This guide discusses the efficacy of various instructional approaches and techniques that teachers and students might not be aware of, such as the benefits of spacing learning, effective study time allocation, and the benefits of testing for subsequent retention.

In terms of improving students' metacognition, teachers should warn students about metacognitive illusions (i.e., systematic errors in metacognitive monitoring) they might encounter when studying specific types of materials or making specific types of judgments (e.g., the foresight bias discussed later). Although students can learn to overcome metacognitive illusions through experience with the illusion-causing materials, only explicit, theory-based training for such illusions transfers to new instances of such materials (Koriat & Bjork, 2006). For this reason, students should be explicitly trained to avoid specific cognitive illusions (e.g., illusions of knowing, use of familiarity heuristics) rather than waiting for them to learn from their mistakes. Further empirical work is certainly required here, as simply teaching or warning students about metacognitive illusions is not adequate to ensure that they will not occur. As described by Wilson and Brekke (1994), such "mental contamination" can be difficult to avoid if one does not feel that it is in fact occurring. Not only will students need to accept that they are prone to specific biases, but they must also learn how to appropriately adjust their metacognitive monitoring judgments to avoid them. Sometimes, however, the metacognitive question can be framed in such a way that the student will recognize the possibility that they might experience a metacognitive illusion. For example, students who are asked whether they will *remember* information on a test will have a different response than students who are asked whether they will *forget* information on a test (Finn, 2008). Framing the question in a different way can facilitate the insight that the student's metacognition might be incorrect.

Monitoring

Metacognitive monitoring focuses on the progress of the cognitive process in which the person is engaged. Such monitoring can take the form of explicit judgments as are typically elicited in laboratory studies of metacognition. These judgments can be made in a simple YES/NO fashion, but are often requested on a continuous scale (e.g., 1–7, 0–100). Numerous monitoring judgments have been examined (e.g., feeling-of-knowing judgments or FOKs, ease-of-learning judgments or EOLs, confidence-in-response judgments, etc.), but this chapter will mainly focus on two: (1) judgments of learning (JOLs), which evaluate one's memory, and (2) metacomprehension judgments, which evaluate one's

understanding of text materials. Although some judgments were initially conceptualized as having direct access to cognitive processes (e.g., Hart, 1967), that position was abandoned in light of evidence suggesting that monitoring judgments are inferential in nature (e.g., Vesonder & Voss, 1985). Rather than having direct access to the cognitive process being evaluated, metacognitive monitoring judgments are made inferentially based on cues that are related to that process (Koriat, 1997). For instance, students studying for a test must make an inference about their learning of the information based on aspects of the materials being studied (e.g., if a chapter in a textbook was written unclearly) or about their experience in studying the materials (e.g., if the information seemed easy to understand). These inferences are also informed by metacognitive knowledge about oneself or the task being judged (Koriat, 1997), such as the knowledge that essay tests require in-depth studying or the belief that one is knowledgeable about physics.

Much research on metacognition has focused on the accuracy of monitoring judgments, which has mainly been conceptualized in two ways: calibration and relative accuracy. The calibration of one's judgments refers to a difference score between the mean of one's predictive judgments and one's performance on the task being judged or predicted. Consider hypothetical participants in a laboratory study on metamemory (metacognition about memory). While studying paired associate items (e.g., two unrelated nouns) for the criterion test, the participants make JOLs on a 0–100% scale indicating the percent likelihood that they will correctly remember each item on a test. Suppose that the overall mean of their JOLs is 88%. Assuming that the participants correctly remember a mean of 66% of the items on the test, their calibration score will be +22%, indicating overconfidence. A group of participants in such a study would be said to demonstrate "good" calibration if the overall mean of their JOLs did not significantly differ from the overall mean of their performance scores.

The relative accuracy of one's judgments refers to a measure of how well one's judgments differentiate performance on the cognitive task being judged. This measure is usually calculated by computing a gamma correlation between one's judgments about individual items and performance on those same test items. Like a Pearson correlation, gamma can range from -1.0 to $+1.0$. If a hypothetical participant gives mostly high JOLs to items they will remember on the test and low JOLs to items they will not remember on the test, their gamma correlation will be positive. Doing the opposite will result in a negative correlation. Assigning JOLs to items at random tends to result in a correlation of zero (or a correlation might even be incalculable). Typically, a gamma correlation is calculated for each participant. The mean gamma correlation is then calculated across all of the participants in a group to estimate the relative accuracy of their judgments; if the mean is significantly greater than zero, the group's judgments are said to be above chance.

Metacomprehension research tends to focus on the relative accuracy of metacomprehension judgments rather than on their calibration. In fact, the relative accuracy of metacomprehension judgments is often referred to simply as metacomprehension accuracy. Although metacomprehension judgments typically demonstrate above-chance relative accuracy, it is typically low (i.e., gamma correlations around or less than 0.3, see Maki, 1998b; Weaver, Bryant, & Burns, 1995). Because of the importance of these self-assessments and because the correlations are typically unacceptably low, much research on metacomprehension judgments has focused on ways to make them more accurate.

It should be noted that the two measures of accuracy discussed above are not always in agreement and do not represent the accuracy of the judgments in the same way (Nelson, 1996); judgments might be considered accurate for one of the measures and inaccurate for the other. For example, Koriat, Sheffer, and Ma'ayan (2002) demonstrated the underconfidence-with-practice (UWP) effect in which learners' JOLs were underconfident from

a second study-test trial on, indicating poor calibration. The underconfident judgments in Koriat et al.'s (2002) study, however, demonstrated good relative accuracy that actually increased across trials.

The accuracy of metacognitive judgments would not matter if metacognition were epiphenomenal. As we discuss later, however, monitoring plays a role in the control of the cognitive process being evaluated (Finn, 2008; Metcalfe & Finn, 2008a; Thiede et al., 2003). Because of this connection, the efficacy of that control will often be linked to the accuracy of the monitoring judgments (Thiede, 1999; Thiede et al., 2003).

Heuristics that Can Produce Inaccuracies in Metacognitive Monitoring

Although both JOLs and metacomprehension judgments typically demonstrate above-chance relative accuracy, the calibration and relative accuracy of these judgments is not always impressive (Maki 1998b; Nelson & Dunlosky, 1991; Weaver & Kelemen, 1997). In many situations, the low predictability of these judgments can be attributed to participants basing them on information that is not diagnostic of their future performance. As Metcalfe, Schwartz, and Joaquin (1993) described:

[When] making judgments related to the external world rather than to their internal states and abilities, people use other heuristics or rules of thumb. These heuristics approximate the uncertain quantity indirectly, rather than measuring the quantity itself. Because they do not measure the quantity directly, such heuristics may result in biases and errors. (p. 860)

In other situations (e.g., delayed JOLs, as we discuss later), calibration and relative accuracy can be quite good because participants base their judgments on information that *is* diagnostic of their future performance. This section reviews some heuristics that are sometimes used to inform metacognitive judgments and that can produce errors and illusions in monitoring. This section is not meant to describe *all* of the potential errors and poor heuristics that might arise in metacognitive monitoring, but simply to provide some illustrative examples. It also suggests some ways that the errors associated with these particular cues might materialize in the classroom. As we learn more about heuristics that lead to accurate and inaccurate metacognitive judgments, we can better understand which heuristics aid or hinder metacognitive accuracy.

Familiarity Heuristics

Learners' familiarity with the information being judged can have an influence on their metacognitive judgments. Because this experience of familiarity often results from prior exposures to and learning of the information, it can be diagnostic of a greater likelihood that the information is known (i.e., it often results in accurate judgments). Unfortunately, familiarity can also arise in situations where it is unrepresentative of knowledge. For example, participants in a study by Reder and Ritter (1992) solved difficult arithmetic problems. After each was presented, they had to quickly choose whether to calculate the answer to each problem or recall it from memory (calculation, of course, was the only option the first time a problem was presented). Participants received 50 points for correctly recalling an answer and 5 points for correctly calculating an answer (but only if the selections and responses were made within the designated time limits). Participants were later paid 0.05 cents for each point earned. By manipulating the occurrence of specific numbers in the set of problems, Reder and Ritter manipulated the participants' familiarity

for the numbers present in the problems independent of their memory for specific problems and their solutions. Participants were able to use their familiarity with the numbers—both independently and as whole problems—to quickly decide whether they knew the answers to the problems. This strategy proved helpful when the specific problem had actually been presented in accordance with the participants' familiarity for the numbers, but proved to be faulty when familiar numbers were combined into novel problems—problems for which the participants could not actually recall an answer.

Although familiarity might sometimes be an appropriate basis for metacognitive judgments, in many situations inappropriate use of familiarity during study can produce monitoring errors. For example, many textbooks provide a list of key terms that students should be knowledgeable about and practice questions that they should be able to answer after reading a chapter. A student who bases a judgment of their memory or understanding for the chapter on their familiarity with these terms or the phrases in the questions rather than on an attempt to define the terms or answer the questions would be likely to overestimate their learning and might be expected to have poor relative accuracy. Such a situation was illustrated well by Metcalfe, Schwartz, and Joaquin (1993), who manipulated the familiarity of cues for feeling of knowing judgments (aka FOKs, which are judgments that one will be able to recognize the correct target word of an incorrectly-recalled cue-target pair at a later time) independent of the retrievability of the targets. This was achieved by manipulating the frequency of the cue words across pairs using a classic interference design (i.e., A-B A-B; A-B' A-B; A-D A-B; and C-D A-B pairs). So, for the A-B A-B procedure, participants studied the same cue-target pairs on each of two lists, whereas for the A-D A-B procedure, participants studied the same cues on each of two lists, but they were matched with different targets on each list (hence a change from D to B). Participants gave higher FOKs to items that had familiar cues (e.g., A-B A-B; A-B' A-B; and A-D A-B pairs) than to pairs that had less familiar cues (C-D A-B) even though interference was produced in some of the former conditions, which impaired recognition performance.

Familiarity can also cause monitoring problems when learning must be monitored across a variety of topics, as is common in educational settings (e.g., studying for multiple final examinations during the same week). Students vary in their knowledge and understanding of the different topics they encounter in school and this information—their domain familiarity—is sometimes used to inform students' metacognitive judgments. Participants in a study by Glenberg, Sanocki, Epstein, and Morris (1987) read and judged their understanding for texts on a number of topics before being tested on their understanding of the texts. Glenberg et al. (1987) found that participants used their familiarity with the topic of each text as the basis for their metacomprehension judgments rather than basing their judgments on their understanding of each particular text. Basing judgments on domain familiarity rather than an understanding of the texts impaired the accuracy of the participants' judgments. Students study topics in multiple domains at the same time in the classroom, and so they might fall prey to a similar error. For example, a student studying for both a mathematics and social studies test the following week might devote more time to studying mathematics because he or she feels less knowledgeable about this topic than social studies. This might often be a good strategy, but if in fact the student has poor understanding for the current social studies lesson and does not factor this into their plan of study, they might go into the social studies test under-prepared.

Fluency Heuristics

Several studies have demonstrated that judgments of learning are sometimes made based on retrieval fluency—the speed and probability with which information is brought to mind

(Benjamin, Bjork, & Schwartz, 1998). Retrieval fluency is predictive of memory performance in many situations (e.g., Serra & Dunlosky, 2005), so using it to inform metacognitive judgments will often be appropriate. It has recently been suggested that the utility of cues such as retrieval fluency and ease of learning are automatically incorporated into metacognitive judgments based on their ecological validity at the time the judgment is being made (Koriat, 2008). For example, Koriat (2008) recently described the easily learned, easily remembered (ELER) heuristic, which stems from the observation that items that are easily learned are typically remembered better than items that are difficult to learn. In the studies reported by Koriat (2008), participants apparently used ease of learning as a cue to inform their metacognitive judgments in a way that was directly correlated with the validity of that cue.

The use of the fluency heuristic as a basis for metamemory judgments in situations in which it is not diagnostic of performance, however, can produce illusions in metacognitive monitoring. For example, Benjamin et al. (1998) purposely chose a task for which retrieval fluency would not be predictive of later test performance. Specifically, participants answered trivia questions (e.g., Who was the second president of the United States?) and then were asked if they could freely recall their answer to the questions, in the absence of the original questions, sometime later. Although participants judged the easiest recalled answers as being the ones they would be the most likely to free-recall later, the most difficult to recall answers were in fact most likely to be free-recalled, presumably because more effortful retrieval attempts were more memorable than less effortful attempts. A crucial part of the free-recall task was recalling the cues themselves; the cues on which the person spent the most time (and hence were most strongly encoded) were those with difficult to retrieve targets. As a result, participants' reliance on retrieval fluency as a basis for their judgments resulted in the judgments being negatively correlated with criterion test performance.

There are also circumstances in which metacognitive judgment accuracy can be impaired by using the fluency of item processing as the basis for the judgments. Interestingly, though, warning people about the possibility of bias can help to offset it. In a study by Jacoby and Whitehouse (1989), participants first studied a list of words. New words were then added to the list and all of the words were presented to the participants, who had to identify each as being either "old" or "new." Prior to this recognition test, some of the new items were primed. Half of the participants were warned about the prime and half were not. Although priming new items increased the processing fluency of the primed items compared to the unprimed items (regardless of the participants' warning condition), the warning did have an effect on whether the items were perceived as new or old. Participants who did not receive the warning were more likely than the warned participants to judge the new, primed items to be old items. Participants who were warned about the prime were more likely to discount the increased fluency of the new, primed items and to correctly identify them as new items. This study demonstrates that some metacognitive errors can be avoided if the potential error is known about beforehand.

Current-Knowledge Heuristic

Sometimes, after information has been obtained or understood, people think they knew or understood it all along. Participants in a study by Fischhoff (1975) read passages detailing outcomes associated with historical events (such as a battle). They then judged if they would have predicted the outcome before reading the passage. Fischhoff demonstrated that these participants could not avoid using their knowledge of the outcome when making this judgment; they even judged that they would have correctly predicted highly unlikely

events (some of which were false). The participants in his study demonstrated hindsight bias—a tendency to use new knowledge when thinking about the past—without even knowing that they were doing so. More so than students, *teachers* should be wary of a form of hindsight bias when judging if their students understand the lesson being taught. Teachers might overestimate their students' understanding of a lesson because—to the teacher—the lesson is easy to understand. Teachers should favor objective measures of their students' understanding (such as quizzes) instead of relying on their own or their students' subjective assessments of the students' learning.

Association Heuristic

Some information seems easier to understand or remember when it is studied than it will actually be to remember or apply later on a test. Koriat and Bjork (2006) termed such an illusion foresight bias and demonstrated a type of paired-associate that produces such an effect. These pairs were composed of two words that had a strong backwards association but a weak forward association (i.e., one would be likely to think of the first word when shown the second word but not likely to think of the second word when shown the first word). For example, consider the pair “fire–blaze.” The word “blaze” is almost always freely associated to the word “fire,” but “fire” rarely—if ever—is freely-associated to the word “blaze.” When such pairs are studied and judged in a typical metamemory procedure, the presence of both words at study makes them seem highly related. At test, however, the stimulus word (fire) is not actually likely to produce the response word (blaze). The association strength present at study produces the illusion that the response word will easily be recalled at test. Such an illusion might occur with more naturalistic materials such as when students learn foreign language vocabulary. An example would be how the apparent similarity between some English words and their Spanish translations (e.g., “computer” and “computadora”) might cause students to experience a foresight bias when studying the words. Although the two words appear similar, this similarity might not be predictive of memory for the specific Spanish translation when only the English word is shown.

Heuristics that Can Cause Illusions of Knowing

One's experience with learning materials sometimes causes the illusion that the materials have been understood when in fact they have not. Participants in a study by Glenberg, Wilkinson, and Epstein (1982) demonstrated an illusion of knowing (i.e., their judgments were overconfident) when asked to rate their comprehension for texts containing factual contradictions. Participants often failed to find these contradictions yet rated their understanding of the texts as being high. This even occurred when factual contradictions were in two adjacent sentences (Glenberg et al., 1982). These findings suggest that readers do not attempt to monitor their understanding across a whole text, but rather at lower levels such as at the per-sentence level. Although educational materials would not be as likely as Glenberg et al.'s (1982) materials to contain factual contradictions, the findings nevertheless suggest that readers might also fail to detect contradictions in their own interpretations of text materials. They might experience that they have understood each sentence well independently, but fail to detect that they have failed to understand the text as a whole, or that their own interpretations of two related sentences are in conflict. As we will discuss later, methods that enhance metacomprehension accuracy tend to focus the reader on their overall understanding of the materials (e.g., Thiede & Anderson, 2003), which is what seems to have been lacking in this example.

Participants in a study by Dunlosky, Rawson, and Middleton (2005) also made meta-comprehension judgments that demonstrated an illusion of knowing. After reading the texts they would be tested on, the participants attempted to answer several questions testing specific pieces of information from each text. They rated their understanding for each piece of information after attempting to answer each question. The participants then completed a criterion test directed at the same information. Judgments made for correct pre-judgment responses tended to be underconfident. More importantly, judgments made for pre-judgment commission errors and partially correct responses demonstrated an illusion of knowing; they were overconfident. The authors interpreted their findings as suggesting that such meta-comprehension judgments reflect the quantity and quality of information recalled during the pre-judgment recall attempt. Without feedback that some of their answers were incorrect, the students judged any answer they gave (even incorrect ones) to be indicative of their understanding for the critical information (Rawson & Dunlosky, 2007).

A form of illusion of knowing can also affect JOLs, particularly when they are made immediately after study (Nelson & Dunlosky, 1991). Consider a participant who is asked to study several paired associates and judge the likelihood that they will remember each item on a test 15 minutes later. If the judgment is to be made immediately after each item is studied, participants will be likely to experience an illusion of knowing because the item will still be in working memory and it will be difficult to evaluate the memorability of the item if it is so strongly represented (Nelson & Dunlosky, 1991, 1992; Nelson, Narens, & Dunlosky, 2004). We will discuss this issue further in the next section when we discuss a way to circumvent this specific problem (i.e., by delaying JOLs).

Heuristics that Can Improve the Accuracy of Metacognitive Monitoring

Given the importance of accurate monitoring in controlling study and the numerous factors that might cause monitoring to be less than optimal, it should not be surprising that much research in the past decade has focused on ways to improve the accuracy of meta-cognitive judgments. In this section we review some factors and methods that improve the accuracy of JOLs and meta-comprehension judgments and discuss research that has sought to determine why accuracy is improved under these conditions.

Debiasing Incorrect Heuristics

As described in the previous section of this chapter, metacognitive judgments are prone to errors, biases, and metacognitive illusions. Experience with and information about these illusions can help to reduce some of these biases. Koriat and Bjork (2006) described one such illusion—foresight bias—in which some to-be-studied pairs of words have a strong backwards association but a weak forward association (e.g., fire–blaze). The presence of both words at study but not at test produced overconfident JOLs for these items. As in King, Zechmeister, and Shaughnessy (1980), Koriat and Bjork (2006) demonstrated that study-test practice reduced this bias, but it did not transfer to new items. Explicit training about the foresight bias and the type of item that causes it, however, not only reduced the bias but also transferred to new items.

Retrieval-Attempt Heuristic

As noted earlier, JOLs often show biased calibration (i.e., overconfidence or underconfidence) and low (but above-chance) relative accuracy. This generalization, however, is

limited to immediate judgments of learning—JOLs made immediately after the study of each of the items for which memory is being judged. Almost two decades ago, Nelson and Dunlosky (1991) demonstrated that delaying the making of JOLs from the study of the corresponding item (by interjecting the study and judgment of other items) greatly improved the accuracy of the JOLs. These delayed JOLs showed significantly greater relative accuracy than immediate JOLs and yielded better calibration curves (which compare JOLs to performance over a range of performance levels) than immediate JOLs did. This effect, commonly referred to as the “delayed-JOL effect,” has been replicated numerous times since the publication of the original article using different materials and procedures. It has been found to effectively improve the accuracy of JOLs made by children (e.g., Schneider, Vise, Lockl, & Nelson, 2000), older adults (e.g., Connor, Dunlosky, & Hertzog, 1997), and special populations such as patients with schizophrenia (e.g., Bacon, Izaute, & Danion, 2007) and attention-deficit/hyperactivity disorder (e.g., Knouse, Paradise, & Dunlosky, 2006).

Despite the robustness and reliability of the delayed-JOL effect, the cause of the effect has been debated since its discovery. In their original article, Nelson and Dunlosky (1991; see also Nelson & Dunlosky, 1992; Nelson et al., 2004) proposed the Monitoring-Dual-Memories (MDM) explanation for the effect. This explanation centers on the traditional distinction between the short-term and long-term memory systems. According to the MDM hypothesis, to make an accurate JOL one must predict the long-term retention of the stimulus-response association. When an immediate JOL is made for a paired associate, however, the stimulus and response word of the pair being judged are both in short-term memory. The presence of the full pair in short-term memory will therefore interfere with the making of accurate immediate JOLs because they assess both short- and long-term memory. In contrast, a delayed JOL is made after the initial presentation of the word pair has left short-term memory. Delayed JOLs only assess one’s long-term memory for an item, which is why they are highly predictive of long-term retention. This also explains why prompting delayed JOLs with the full word pair instead of the stimulus alone can reduce the accuracy of the judgments (Dunlosky & Nelson, 1992).

Spellman and Bjork (1992) countered the MDM explanation for the delayed-JOL effect by suggesting that delaying JOLs improved memory for the high JOL items but did not affect metamemory. Their “self-fulfilling prophecy” explanation also assumed that the participants attempt to retrieve the response word when making a delayed JOL. Items that are not recalled receive low JOLs while those that are recalled receive high JOLs. Although up to this point this explanation is not entirely different from the MDM explanation, Spellman and Bjork pointed out that items that *were* successfully retrieved therefore received retrieval practice, making them more likely to be recalled later. The delayed JOLs create a self-fulfilling prophecy by increasing the likelihood that the successfully retrieved items will also be retrieved later. Kimball and Metcalfe (2003) later provided evidence to support Spellman and Bjork’s (1992) explanation for the effect. Specifically, Kimball and Metcalfe showed that re-exposing the response word after a stimulus-prompted delayed JOL was made reduced the relative accuracy of the judgments. This occurred because memory for the low-JOL items was increased by this exposure while memory for the high-JOL items was not affected. Such re-exposure did not affect the relative accuracy or recall of stimulus-prompted immediate JOLs or stimulus-response-prompted delayed JOLs. This evidence provides support for the idea that delayed JOLs are accurate because making such judgments affects memory for the items being judged. The MDM explanation for the delayed-JOL effect, however, cannot explain the effects of re-exposure without positing an additional mechanism (Kimball & Metcalfe, 2003).

In response to Spellman and Bjork (1992), Nelson and Dunlosky (1992; Nelson et al.,

2004) provided evidence to show that delaying JOLs improved the metacognitive accuracy of the judgments, not memory for the items. Specifically, they pointed out that mean recall was the same for items slated with either immediate or delayed JOLs in the Nelson and Dunlosky (1991) dataset, which suggests that delaying JOLs does not boost memory for the items being judged. As Kimball and Metcalfe (2003) pointed out, though, only those items that were correctly retrieved at the delay would be expected to get a boost in memorability. Memorability for those items that were *not* correctly recalled at the delay would not be boosted. In contrast, memory for these items might actually be impaired (relative to the immediate-judgment items) because, in the former but not the latter case, they would be rehearsed (if in a massed way). Thus, although overall recall levels in the immediate and delayed JOL cases do not speak to Spellman and Bjork's hypothesis, the recall of retrieved versus unretrieved items does. Using explicit pre-judgment recall attempts, Nelson et al. (2004) showed that delayed JOLs obtain most of their relative accuracy advantage when participants give high JOLs to successfully retrieved items and low JOLs to items they could not recall. Feedback such as these retrieval attempts can be used to reduce both cognitive and metacognitive errors; results from a non-JOL study by Butler, Karpicke, and Roediger (2008) show that providing students with feedback can increase their recall of low-confidence correct responses, whereas memory for these items will be hampered if feedback is not given.

More recent research by Son and Metcalfe (2005) demonstrated that some low JOLs are made rather quickly—too quickly for retrieval to be attempted. This suggests that JOLs are made in two stages, the first being a quick evaluation of whether one recognizes the cue or not and the second being a target retrieval attempt that can be used to inform the JOL further (see also Metcalfe & Finn, 2008b). This might help to explain one criticism of the memory-based explanation for the delayed-JOL effect, which is that delayed-JOL items do not consistently show a recall advantage over items for which immediate (or no) JOLs were made. If all delayed JOLs were made based on a retrieval attempt, delayed-JOL items would be expected to consistently show a recall advantage over immediate-JOL items because, as long as target retrieval is successful, of course, the pre-JOL retrieval should enhance performance on the criterion test. Metcalfe and Finn (2008b) had participants make either speeded (less than 0.75 seconds to respond) or unspeeded (unlimited time to respond) delayed JOLs. Memory was enhanced for the unspeeded JOLs relative to the speeded JOLs, presumably because participants in the latter case did not have time to make a retrieval attempt (and instead had to base their JOLs primarily on the familiarity of the cue).

Although the explanation for why the delayed-JOL effect occurs is still under debate, there is consensus on the reliability of the finding itself: empirical evidence has consistently indicated that explicit retrieval attempts help to make delayed JOLs accurate (Finn & Metcalfe, 2007, 2008; Kimball & Metcalfe, 2003; Nelson et al., 2004; Spellman & Bjork, 1992). For this reason, students should be instructed to attempt retrieval and use the outcome as the basis for delayed JOLs. Such retrieval attempts are highly predictive of later test performance (Finn & Metcalfe, 2007; Nelson et al., 2004), which explains why basing delayed JOLs on their outcome predicts test performance so well. As suggested by Thiede (1999), “a new focus for metacognitive training may be to teach students to discriminate between what they know versus what they don't know” (pp. 666–667). Having students make delayed JOLs is one way to do this. Making such delayed retrieval attempts also has the secondary advantage of boosting retention performance by serving as a spaced practice trial (Kimball & Metcalfe, 2003; Spellman & Bjork, 1992). It is also advisable to look up the answers to items given low delayed JOLs because corrective feedback will enhance the recall of these items (Butler et al., 2008).

Memory for Past Test Heuristic

Providing students with retrieval opportunities prior to making JOLs increases the accuracy of those JOLs (King et al., 1980; but see Koriat et al., 2002). The effect does not, however, transfer to new items (King et al., 1980). This suggests that although the outcome of a retrieval attempt can be useful for informing JOLs (Finn & Metcalfe, 2007; Nelson et al., 2004), this occurs because it provides information about memory for the item retrieved, *not* information about the retrieval task itself. Recently, Finn and Metcalfe (2007, 2008) demonstrated the Memory for Past Test (MPT) Heuristic. These studies demonstrate that JOLs are based—at least in part—on earlier retrieval outcomes. More specifically, Finn and Metcalfe (2007, 2008) considered whether the MPT heuristic might be responsible for both the decreased calibration and increased relative accuracy associated with the UWP effect (Koriat et al., 2002). They hypothesized that participants were, in part, basing their Trial 2 immediate JOLs on whether or not an item was recalled on Trial 1. Although this contributed to the underconfidence of their Trial 2 JOLs (because some of the items that were incorrect on Trial 1 had been learned on Trial 2), it also contributed to their increased relative accuracy on Trial 2 (because, overall, their Trial 1 recall performance did predict their Trial 2 performance). Although the MPT heuristic can contribute to calibration errors (i.e., underconfidence with practice) it can also increase relative accuracy. Because good relative accuracy is important for the control of study (Thiede, 1999) and there in fact might be some advantage to being slightly underconfident when studying, instructing students to consider their past test performance seems like a simple and effective way to improve the accuracy of their memory monitoring.

Summarization Heuristic

Asking learners to summarize what they have read before judging how well they understand it is one way to improve the accuracy of metacomprehension judgments (Thiede & Anderson, 2003). Summarization seems to improve the metacomprehension accuracy of all readers, including high-ability readers (Griffin, Wiley, & Thiede, 2008). This improvement, however, is only found when the writing of the summaries and making of the judgments occur at a delay from the reading of the texts (Thiede & Anderson, 2003). Thiede and Anderson (2003) evaluated two potential explanations for this effect. Both involved the participants' use of aspects of the summaries as cues for their judgments, but neither was consistently supported. Despite this earlier conclusion, Griffin et al. (2008) later concluded that summarizing improves metacomprehension accuracy “by increasing the access and salience of cues that are actually predictive of performance that requires inference and situation-model level understanding” (p. 102), but they did not examine this possibility directly. The difference between these two conclusions is that Thiede and Anderson (2003) focused on attributes of the summaries written, such as their length, without considering how indicative of higher-level understanding such cues are. Future research should attempt to determine what information about the summaries is used as cues when summarizing improves the metacomprehension accuracy of metacomprehension judgments.

Although summarization helps students make more accurate metacomprehension judgments, writing summaries can be a time-consuming endeavor. For this reason, researchers have sought out more time-efficient methods to improve the accuracy of these judgments. One such alternative involves a direct analogue to writing summaries: delayed-keyword generation. Having students generate keywords (important points from the text) a short

time after reading a text but before making a metacomprehension judgment for it makes the judgments more accurate but requires less time than would be required to write a full summary (Thiede, Anderson, & Therriault, 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005). According to Thiede et al. (2005), delayed-keyword generation improves the accuracy of the judgments by providing the learner with cues that are diagnostic of their understanding of the text. As long as the test will tap the learner's understanding of the text, these cues will also be predictive of performance on that test. Like summarization, keyword generation only improves the accuracy of metacomprehension judgments when employed at a delay from the reading of the texts. Simply delaying metacomprehension judgments, however, does not improve their accuracy (Dunlosky et al., 2005; Maki, 1998a); the delayed judgments must be accompanied by cues that are diagnostic of the learners' understanding of the text (Griffin et al., 2008; Thiede et al., 2005). Summarization and keyword generation are two ways to provide such diagnostic cues.

Knowledge of Test Heuristic

Letting students know how information will be tested should help them judge their learning more accurately and study more efficiently. Thomas and McDaniel (2007) recently examined how the match between the type of encoding and type of testing affected metacomprehension accuracy. They demonstrated that participants made more accurate metacomprehension judgments when encoding and testing either were both detail-focused or were both concept-focused, which also allowed the students to more effectively control their further study of the materials. These results suggest that if students can be trained to study in a way that is conducive to the type of test they expect, they can improve the accuracy of their metacomprehension judgments while studying. Informing students of the format of the test should also help them to set an optimal "grain size" for their metacomprehension judgments and further increase their accuracy (Dunlosky, Rawson, & Middleton, 2005). Of course, this suggestion also involves teaching students which types of encoding are most appropriate for certain types of test (i.e., increasing their metacognitive knowledge).

Another method for improving the accuracy of metacomprehension judgments is to have students make term-specific metacomprehension judgments (Dunlosky, Rawson, & Middleton, 2005). In contrast to typical metacomprehension judgments, which are made for an entire text or for large portions of a text, term-specific judgments are made for specific pieces of information from the text pre-selected by the experimenter (or, presumably, the teacher). Rather than judging how well they understand a text, students making term-specific judgments judge their ability to answer specific questions that are very similar to those they will actually be tested with later. These judgments demonstrate greater metacomprehension accuracy than typical metacomprehension judgments, especially when participants are forced to attempt to retrieve answers for each term-specific judgment (Dunlosky, Rawson, & Middleton, 2005) and when feedback is given on the correctness of their answers (Rawson & Dunlosky, 2007). Although this method of improving the accuracy of students' metacomprehension judgments is promising for situations where learners' understanding of text will be tested via retention questions, to date its efficacy has not been evaluated when the test will require a more advanced understanding of the text (i.e., inference or transfer questions). Future research should explore the generality of this method to other question types.

Control

Metacognitive control can be defined as any instance of cognitive control that is informed by metacognitive knowledge or monitoring. Concerning studying specifically, control might involve the choice of which items to study and the allocation of study time (e.g., devoting more study time to certain pieces of information than others) or strategy selection (e.g., choosing what is believed to be an effective way to study certain materials; changing a study strategy when an earlier choice proves inefficient). Regardless of which form(s) of control is used while studying, that control will be informed by the student's metacognitive knowledge and monitoring and will be dependent upon the accuracy of both.

Types of Control

Allocating Study Time

The general design of laboratory studies that have examined the relationship between monitoring and study-time allocation involves participants studying items or materials for a fixed period of time, judging their learning for those materials, and then re-studying the materials for the amount of time the participants feel is appropriate. In some cases, the mean amount of time spent studying different types of items is compared (e.g., time spent studying easy versus difficult items). Alternatively, a correlation can be calculated between the magnitude of the participants' JOLs and their continued study time. The correlation between JOLs and study time is usually negative (for a review see Son & Metcalfe, 2000), which also obtains for other judgments (i.e., EOLs and FOKs, see Nelson & Leonesio, 1988). In other words, students tend to devote more study time to those items or materials on which they judge their learning to be low than to those items they believe to be known.

This finding supports the idea that students study to reach a goal state in which to-be-learned information is learned to a pre-set criterion level (Dunlosky & Thiede, 1998). More specifically, the discrepancy-reduction hypothesis forwarded by Dunlosky and Thiede (1998) predicts that students will devote study time to the to-be-learned information until the discrepancy between its JOL and the goal state is eliminated. This account explains the negative correlation typically found between JOLs and study time. One shortcoming of the discrepancy-reduction model, however, is that it does not contain a formal stopping mechanism, or some way for the student to stop studying a piece of information they either feel they are not making progress towards learning or simply cannot ever learn. In other words, the model assumes that continued studying allows students to learn any piece of information to criterion. This is of course unlikely; some information will undoubtedly be too difficult for every student to master and continued studying of the information on its own will likely be in vain (Nelson & Leonesio, 1988). Although Dunlosky and Thiede's (1998) model did not contain a formal mechanism to explain how students decide to cease study, they did suggest one that centered on the student's perceived *rate* of learning. An almost identical mechanism was later proposed and evaluated by Metcalfe and Kornell (2005), who demonstrated that as learners study items over time, their JOLs for the items do not continue to increase. As the rate of perceived learning decreases, students are more likely to cease studying.

Although the negative relationship between JOLs and study time has been demonstrated under varied conditions, neither it nor the discrepancy-reduction hypothesis fully captures the relationship between monitoring and the allocation of study time in all situations. For instance, following Thiede and Dunlosky's (1999) seminal study, Son and Metcalfe (2000) showed that study time is not always allocated preferentially to difficult items; factors such

as time pressure and learning goals play a role in the choices made. Participants in Son and Metcalfe's (2000) first experiment were given insufficient time to study all of the items presented, so they devoted study time to the easier items. The discrepancy-reduction model would not predict this pattern of results. When participants in Son and Metcalfe's second experiment studied shorter materials, they devoted more of their time to studying more difficult items. In other words, the pattern of study-time allocation predicted by the discrepancy-reduction model only obtained when there was enough time given for difficult items to be studied.

Thiede and Dunlosky (1999) examined the role of their discrepancy-reduction model in both the allocation of study time *and* study choice (selecting which items to study). More difficult items require more learning to reach a goal-based learning state, so the discrepancy-reduction model predicts that these items should be studied for a longer amount of time than easier items (as in Dunlosky & Thiede, 1998). The model also predicts that more difficult items should also be selected for study *more often* than easier items, but this did not obtain. Instead, Thiede and Dunlosky obtained a positive correlation between participants' JOLs and study choices. In other words, the participants chose to re-study items they assigned higher JOLs to. To account for this result, Thiede and Dunlosky (1999) proposed a multi-level system capable of meeting learning goals in an efficient way. For example, if the learning goal was to learn all items to a criterion level, the original discrepancy-reduction approach would *have* to be used and more difficult items would be studied longer than easier items, resulting in the typical negative correlation between JOLs and study time. If a less ambitious goal was to be met, such as learning 50% of all items to a criterion, the planning system would choose the easiest 50% of the items to learn, explaining positive correlations between JOLs and study choice in such situations. Within this subset of selected items, though, JOLs and study time would still be negatively correlated as predicted by the original instantiation of the discrepancy-reduction model.

Metcalfe and Kornell (2003) provided further support for the idea that study time is strategically allocated to items. They presented three single-word Spanish-English translations to participants at a time. One of the three translations was easy, one was of medium difficulty, and one was difficult. Participants in their study chose to study the easiest items first before studying the medium-difficulty and difficult items if they had study time remaining. They devoted more study time to the medium-difficulty items than to either the easy or difficult items. This pattern of study proved to be efficient; easy items benefited greatly from short amounts of study time, whereas medium-difficulty items required more study time to demonstrate learning gains. More importantly, the pattern of results obtained by Metcalfe and Kornell (2003) supported an alternative to the discrepancy-reduction model of study-time allocation, the region of proximal learning (Metcalfe, 2002). The idea behind the region of proximal learning model is that there is some set of to-be-learned information that is just beyond the current understanding of the learner that will therefore most benefit from further study. This is why both expert and novice adult participants in the studies conducted by Metcalfe (2002) preferentially allocated their study time to items that would most benefit from continued study. Novices (participants who were not experts in Spanish) devoted most of their study time to the easy and medium-difficulty items, whereas experts (participants who declared themselves to be expert Spanish speakers and who scored above 30% on the criterion test) devoted most of their time to the difficult items.

Based on the test performance of these two groups, their strategies were logical. For the novices, the easy and medium-difficulty items benefited from the study time that was allocated to them. Although the experts did not devote much study time to these same items, they nevertheless answered many of them correctly on the test, indicating that they

were correct by not devoting study time to them. Novice and expert children in Metcalfe's (2002) studies devoted study time similarly to the adults; novice children devoted most of their study time to easy items while expert children devoted less of their time to the easiest items. Although this suggests that children can be just as strategic in allocating study time as adults, the study-time allocation of the expert children was not as effective as that of the expert adults.

It is important to note that both groups of novices would be considered to have demonstrated inefficient study-time allocation based on the discrepancy-reduction model because they did not devote their study time preferentially to the difficult items. Based on the region of proximal learning idea and the data presented by Metcalfe (2002), though, this strategy proved to be beneficial because allocating study time to the easier items would result in greater overall learning than devoting it to difficult items would. Metcalfe and Kornell (2005) further delineated how the region of proximal learning model affects the allocation of study time by demonstrating that it involves two components: choice (i.e., study choice) and perseverance (i.e., study-time allocation). The first component, choice, has two stages: a yes/no decision to study an item and an evaluation of the priority of the item for study. The decision to study or not is based on an attempt to eliminate items that are already known from the study queue. Priority is determined based on the region of proximal learning; items closest to being learned will be studied first. The second component, perseverance, is based on the stopping mechanism described earlier that considers the rate of learning being experienced for an item. Allocating study time in this way has been demonstrated to be efficient for improving learning and is how adults actually do allocate their study time naturally, both of which support the region of proximal learning model (Kornell & Metcalfe, 2006).

Strategy Selection

Sometimes students realize that the way they are currently studying is not producing the desired learning outcomes. In such situations, students might switch (i.e., “toggle”) or modify (i.e., “edit”) their current strategy in favor of a different approach (Winne & Hadwin, 1998; see also Siegler, 2002, for a similar concept). This is another way that metacognition—both knowledge and monitoring—can affect the control of study. In fact, Winne and Hadwin (1998) described editing and toggling as a more advanced form of metacognitive control than self-regulated study. They also cautioned that when a task is complex and metacognitive knowledge about it is low, metacognition could actually hurt performance on the task. In such cases they recommend that the learner be given explicit sub-goals to help scaffold their control. Although young children can make decisions concerning their study choices, they are not as adept at doing so as adults are. For example, Son (2005) recently demonstrated that children could make short-term study choices that optimized their learning, but had difficulty doing so when the decision required thinking about long-term future outcomes (e.g., to space practice instead of massing practice).

Monitoring Accuracy and Control

If metacognitive monitoring is used to control study, it seems logical that the control of study is most effective when monitoring is accurate.* Early studies examining the link

* Although we have stressed making *accurate* metacognitive judgments in this chapter, there might be situations where it is advantageous to make inaccurate judgments. In particular, there might be advantages to being

between the accuracy of metacognitive monitoring and learning often did not allow students an opportunity to use their monitoring to affect their learning. Typically, then, these studies failed to show a relationship between the two (for a review see Thiede, 1999). It was largely concluded that monitoring accuracy does not affect learning. In contrast, Thiede (1999) allowed students to use their monitoring to affect their learning. Participants studied 36 single-word Swahili–English translations for four seconds each on a computer. After the participants studied the entire list, they made JOLs for each item and then were tested on their memory for the translations. After this test, participants saw all 36 items in an array and could choose to re-study as many of the 36 items as they would like. This procedure was repeated either for six trials or until a participant learned all 36 items. Regression analyses conducted by Thiede (1999) demonstrated that better test performance was related to more accurate monitoring (relative accuracy) and better self-regulated learning (choosing lesser-known items for re-study). Importantly, better test performance was not related to selecting more items for re-study, which further supported the conclusion that more accurate monitoring led to more efficient study.

The accuracy of metacognitive monitoring judgments also affects the self-regulated study of text materials. Participants in a study by Thiede et al. (2003) studied several expository texts and made a metacomprehension judgment for each. The accuracy of their judgments was manipulated between-participants by having the participants generate keywords (important words from the text that capture its main idea) either immediately after reading each text, after reading all of the texts (delayed), or not at all. Participants then answered detail- and inference-based questions for each text and, much as in Thiede (1999), were given an opportunity to select texts for re-study. Participants in the three groups performed equally on the first test, but participants who generated keywords at a delay from the reading of the texts achieved higher relative accuracy than participants in the other two conditions (i.e., a gamma correlation of about +0.7 compared to about +0.4 for the no keyword group and about +0.3 for the immediate keyword group). After studying the selected texts again, participants answered new detail and inference questions as well as the same questions answered earlier. Although participants in the three groups chose to re-study the same number of texts, the groups differed in which texts they chose and how well they performed on the second test. Participants in the delayed-keyword generation group chose to re-study texts that they did not understand well after the first study attempt, whereas participants in the other two groups chose to re-study texts that they understood moderately well. Participants in the delayed-keyword group also performed better than participants in the other two groups on the new questions. As a whole, these results suggest that greater relative accuracy of metacomprehension judgments is associated with better self-regulated learning (choosing to re-study the least-understood texts) and better learning.

The accuracy-control link is apparent in the case of relative accuracy (i.e., Thiede, 1999; Thiede et al., 2003). Poor calibration—in the form of an illusion of knowing

underconfident when making learning judgments. Underconfident judgments should result in extra study, and might be appropriate in real-life studying situations where the test is typically delayed from the study episode (as opposed to experimental situations where the test typically occurs shortly after the study episode). It is difficult to imagine situations in which making judgments with poor relative accuracy or overconfident judgments might be advantageous for adults (young or old). For children, however, it has been argued (e.g., Bjorklund & Green, 1992) that overconfidence can be advantageous because their cognitive abilities and knowledge are still developing. If children accurately knew that they would perform poorly on most things they attempted to do (i.e., if their self-efficacy matched their abilities), they might not attempt to do those things, and therefore might never learn how to do them.

(overconfidence) or of not knowing (underconfidence)—has also been demonstrated to have an effect on control processes in that over- and underconfidence can spuriously alter whether or not people choose to study. When students think they have learned the study materials—regardless of whether they really have or not—they will choose not to study; when students think they have not learned the study materials—again, regardless of whether they really have or not—they will choose to study. For example, participants in the first experiment presented by Metcalfe and Finn (2008a) studied paired associates on two study-judge-test trials. Half of the items were presented once on Trial 1 and three times on Trial 2 (“1-3” items). The other items were presented three times on Trial 1 and once on Trial 2 (“3-1” items). JOLs were made for all items on both trials. Participants were given the opportunity to select items for re-study after making each Trial 2 JOL but re-study did not actually occur. Although recall was greater for the 3-1 items on Trial 1, it was equated on Trial 2. JOLs for the 3-1 items, however, were greater than those for the 1-3 items on both trials. Most important, participants chose to re-study 1-3 items more often than 3-1 items. This pattern reflected the participants’ misperception that they knew the 1-3 items less well than they knew the 3-1 items. This effect obtained even though the relative accuracy of the Trial 2 JOLs was equal for both subsets of items. This suggests that distortions of calibration might exert a negative effect on the control of study even in situations where relative accuracy is good (i.e., Koriat et al., 2002; Metcalfe & Finn, 2008a). This possibility deserves further investigation, as little research has examined the effect of poor calibration on study efficiency or study choices.

It is not enough for students to make accurate metacognitive judgments—they also need to know how to transform them into a study plan. One such plan is suggested by Metcalfe’s (2002) Region of Proximal Learning (RPL) model, which suggests that students should allocate study time to the information that is most amenable to learning (Metcalfe, 2002; Metcalfe & Kornell, 2003). As demonstrated by Kornell and Metcalfe (2006), this strategy is effective for enhancing learning, so students should be trained to allocate their study time in this way. As suggested by Metcalfe and Kornell (2005), the first step should be for students to eliminate from study information that they already know. Making accurate metacognitive judgments here will be critical, as students might mistakenly eliminate information that they do not yet know *well*; caution should be exercised when making such decisions. It would be better to re-study known information than to not study at all information that could actually benefit from further study. Once already-known information is eliminated from study, students should determine how well they have already learned the remaining to-be-learned information and devote study time to that information which is most amenable to learning, e.g., to information or items given the highest JOLs. When that information is learned well, it too should be eliminated from study so more difficult information can be attended to. Studying in this way should produce the largest learning gains in a limited amount of time regardless of the ultimate learning goal (i.e., to learn all materials to mastery or to just learn some portion of the information). Loftier goals will likely require more overall study time and better time management, so students might also need to be trained on how to set goals and how to manage their time. Training on goal setting and time management has been shown to improve self-regulated learning behaviors when combined with training on metacognition (Azevedo & Cromley, 2004).

Of course, there is a difference between information that has, for all purposes, been committed to memory (i.e., that can readily and reliably be retrieved from memory) and newly-learned information that might be recalled once but might not be recalled again in the future. As demonstrated by Kornell and Bjork (2008), good students naturally “drop” items from study when using flashcards, but often do so in ineffective ways. For example,

in their experiments some students dropped items that seemed learned (i.e., because they were successfully recalled once) that in fact were not *well* learned. This becomes increasingly common as more items are dropped and the intervals between study trials for individual items decrease (see Pyc & Rawson, 2007, for a demonstration of the effectiveness of dropping items from study and its dependency on study intervals). Other students in Kornell and Bjork's (2008) experiments dropped items that they felt would be too difficult to learn in the given time constraints, which hurt their overall performance because the items might have been learnable if they had studied the items repeatedly. These results suggest that before students should be instructed to "drop" items from study, they must be taught when it is advantageous to drop items and when it is not.

Summary and Future Directions

In this chapter we reviewed metacognition, its role in studying, and how errors in monitoring can reduce the efficacy of study. We also reviewed methods that improve the accuracy of metamemory and metacomprehension judgments and suggested ways for students and teachers to implement them in the classroom and during study. Although we based these suggestions on empirical evidence, further research is needed to ensure that implementing these methods as we suggest is effective. Research should also continue to find even more effective ways of improving metacognitive knowledge, monitoring, and control.

Author Note

Michael Serra is now at Texas Tech University.

This research was supported by National Institute of Mental Health grant RO1-MH60637. We thank John Dunlosky for his help and comments.

Correspondence concerning this article should be addressed to Janet Metcalfe, Department of Psychology, Columbia University, 406 Schermerhorn Hall, 1190 Amsterdam Avenue, New York, NY 10027. E-mail: jm348@columbia.edu

References

- Azevedo, R. & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*, 523–535.
- Bacon, E., Izaute, M., & Danion, J. (2007). Preserved memory monitoring but impaired memory control during episodic encoding in patients with schizophrenia. *Journal of the International Neuropsychological Society, 13*, 219–227.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55–68.
- Bjorklund, D. F. & Green, B. L. (1992). The adaptive nature of cognitive immaturity. *American Psychologist, 47*, 46–54.
- Butler, A. C., Karpicke, J. D., & Roediger, H. L., III (2008). Correcting a meta-cognitive error: Feedback enhances retention of low confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 918–928.
- Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12*, 50–71.
- Dunlosky, J. & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*, 374–380.
- Dunlosky, J. & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica, 98*, 37–56.

- Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate monitoring and accessibility hypotheses. *Journal of Memory and Language*, *52*, 551–565.
- Dunlosky, J., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgments about judgments of learning [Special issue]. *Journal of General Psychology*, *132*, 335–346.
- Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition*, *36*, 813–821.
- Finn, B. & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 238–244.
- Finn, B. & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language*, *58*, 19–34.
- Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288–299.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*, 906–911.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, *116*, 119–136.
- Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*, 597–602.
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, *36*, 93–103.
- Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, *6*, 685–691.
- Jacoby, L. L. & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, *118*, 126–135.
- Kimball, D. R. & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition*, *31*, 918–929.
- King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology*, *93*, 329–343.
- Knouse, L. E., Paradise, M. J., & Dunlosky, J. (2006). Does ADHD in adults affect the relative accuracy of metamemory judgments? *Journal of Attention Disorders*, *10*, 160–170.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370.
- Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, *36*, 416–428.
- Koriat, A. & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1133–1145.
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131*, 147–162.
- Kornell, N. & Bjork, R. A. (2008). Optimizing self-regulated study: The benefits—and costs—of dropping flashcards. *Memory*, *16*, 125–136.
- Kornell, N. & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 609–622.
- Maki, R. H. (1998a). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition*, *26*, 959–964.
- Maki, R. H. (1998b). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Mahwah, NJ: Erlbaum.
- Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, *131*, 349–363.

- Metcalfe, J. & Finn, B. (2008a). Evidence that judgments of learning are directly related to study choice. *Psychonomic Bulletin Review*, *15*, 174–179.
- Metcalfe, J. & Finn, B. (2008b). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 1084–1097.
- Metcalfe, J. & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General*, *132*, 530–542.
- Metcalfe, J. & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*, 463–477.
- Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 851–864.
- Nelson, T. O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. *Applied Cognitive Psychology*, *10*, 257–260.
- Nelson, T. O. & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, *2*, 267–270.
- Nelson, T. O. & Dunlosky, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science*, *3*, 317–318.
- Nelson, T. O. & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 676–686.
- Nelson, T. O. & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: The MIT Press.
- Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment Recall And Monitoring (PRAM). *Psychological Methods*, *9*, 53–69.
- Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing Instruction and Study to Improve Student Learning* (NCER 2007–2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ncer.ed.gov>.
- Pyc, M. A. & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35*, 1917–1927.
- Rawson, K. A. & Dunlosky, J. (2007). Improving students' self-evaluation of learning of key concepts in textbook materials. *European Journal of Cognitive Psychology*, *19*, 559–579.
- Reder, L. M. & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 435–451.
- Schneider, W., Vise, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning (JOL) task. *Cognitive Development*, *15*, 115–134.
- Serra, M. J. & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 1258–1266.
- Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31–58). Cambridge, UK; Cambridge University Press.
- Son, L. K. (2005). Metacognitive control: Children's short-term versus long-term study strategies [Special issue]. *Journal of General Psychology*, *132*, 347–363.
- Son, L. K. & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 204–221.
- Son, L. K. & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage model. *Memory & Cognition*, *33*, 1116–1129.
- Spellman, B. A. & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, *3*, 315–316.

- Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review*, 6, 662–667.
- Thiede, K. W. & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, 28, 129–160.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95, 66–73.
- Thiede, K. W. & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1024–1037.
- Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1267–1280.
- Thomas, A. K. & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition*, 35, 668–678.
- Vesonder, G. & Voss, J. (1985). On the ability to predict one's own responses while learning. *Journal of Memory and Language*, 24, 363–376.
- Weaver, C. A., III, Bryant, D. S., & Burns, K. D. (1995). Comprehension monitoring: Extensions of the Kintsch and van Dijk model. In C. A. Weaver III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 177–193). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Weaver, C. A., III & Kelemen, W. L. (1997). Judgments of learning at delays: Shifts in response patterns or increased metamemory accuracy? *Psychological Science*, 8, 318–321.
- White, B. Y. & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16, 3–118.
- Wilson, T. D. & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, 116, 117–142.
- Winne, P. H. (1995). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences*, 8, 327–353.
- Winne, P. H. & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Erlbaum.

16 Self-Regulation

Where Metacognition and Motivation Intersect

*Barry J. Zimmerman and
Adam R. Moylan*

Graduate School and University Center, City University of New York

“Education is not the filling of a pail but the lighting of a fire”

(W. O. Yeats)

One of the most challenging issues that confronts educational researchers is explaining how students learn in self-regulated contexts, such as when studying or practicing on their own. Acquisition of skills in these demanding contexts requires more than passive compliance with prior directions; it also involves personal initiative, resourcefulness, and persistence—the motivational fire to which Yeats refers. There is evidence that proactive students often seek to create their own enriched environments for learning (Scarr & McCartney, 1983). However, learning in self-regulated contexts can be challenging for students due to (a) competing activities, such as watching television or conversing with friends, (b) insufficient knowledge about how to proceed, (c) difficulty in judging the quality of one’s learning, and (d) insufficient incentives. These attention, retention, self-awareness, and motivation issues have been studied as important attributes of self-regulated learners. Self-regulated learning refers to self-generated thoughts, feelings, and actions for attaining one’s learning goals.

Research on self-regulated learning initially focused most heavily on students’ use of key metacognitive processes, such as strategy use and self-monitoring (Borkowski, 1985; Brown, Bransford, Ferrara, & Campione, 1983). Metacognition refers to knowledge, awareness, and regulation of one’s thinking. There is a growing literature indicating that a student’s use of these metacognitive processes to learn is not merely a question of competence but is also a question of motivation to explain his or her willingness, effort, and persistence. There is considerable evidence that the effects of strategy training are not well “maintained” over time or “transferred” to non-experimental contexts (Pressley & McCormick, 1995, p. 33). These shortcomings have been attributed to deficiencies in both metacognition and motivation. Regarding metacognitive deficiencies, students may not be aware that a strategy could be used in a new situation. Regarding motivational deficiencies, students may fail to use a known strategy because they did not enjoy carrying it out or did not feel its outcomes were worth the effort (Rabinowitz, Freeman, & Cohen, 1992). These two beliefs—intrinsic value and outcome expectations—are two widely studied sources of motivation to learn on one’s own. There is now an extensive effort to include motivational constructs along with metacognitive processes in models of self-regulated learning (Schunk & Zimmerman, 2007).

In this chapter, we use a social cognitive model of self-regulated learning to explain the intersection of students’ metacognitive processes and motivational beliefs. Then, we describe how these processes and beliefs can be assessed as they occur during learning using a microanalytic methodology. Third, the usefulness of the three-phase model for training

reactive learners to become more proactive is discussed. Fourth, we consider the relation between metacognitive calibration and two key motivational measures: self-efficacy and self-evaluation. Fifth, we describe the results of a recent study that was designed to improve at-risk students' calibration and academic achievement. Finally, we describe the results of two recent meta-analyses of self-regulatory training interventions, with particular attention to the interactive role of metacognitive processes and motivational beliefs.

A Cyclical Phase Model of Self-Regulatory Feedback

A defining feature of self-regulation theories is a personal feedback loop. This loop refers to information regarding one's performance or outcomes that is used to make subsequent adaptations (Hattie & Timperley, 2007). This feedback may be social, such as guidance or praise from a teacher, peer, or a parent. It may also be environmental, such as from a task or the immediate context, or this feedback may be personal, such as awareness of covert mental outcomes, physiological outcomes, or behavioral outcomes. Building on earlier findings, social cognitive researchers' understanding of the nature and cyclical functioning of personal feedback loops has become more detailed and complete over time.

According to a social cognitive model of self-regulation, students' feedback loops involve three cyclical phases (see Figure 16.1). The forethought phase refers to learning

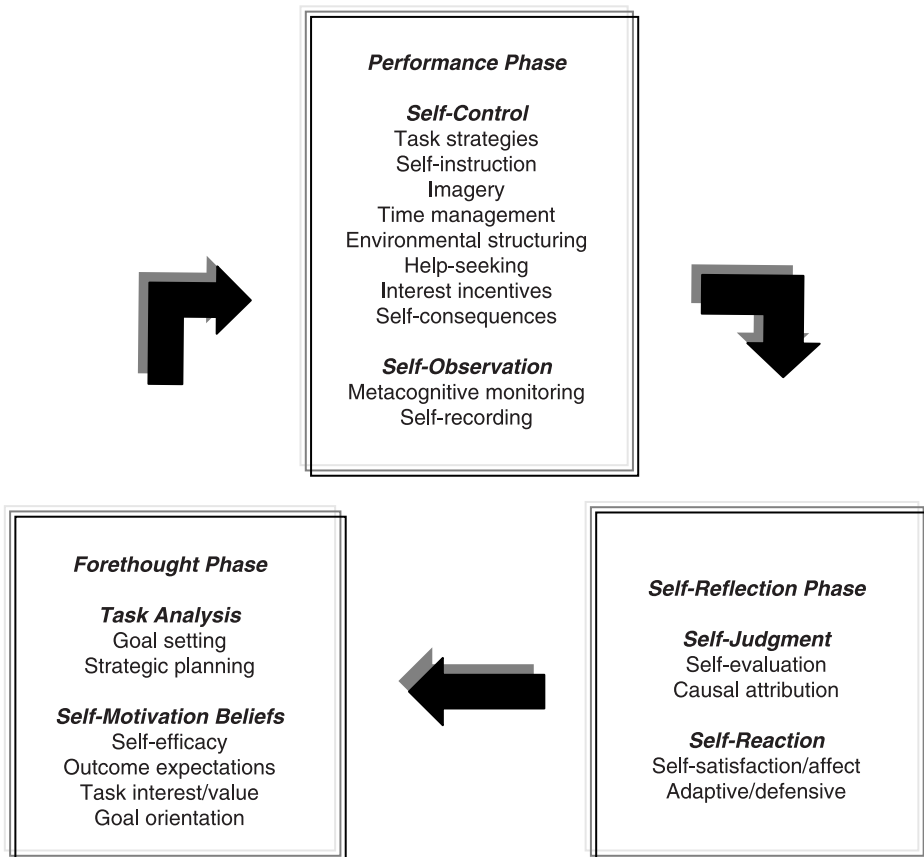


Figure 16.1 A cyclical phase model of self-regulation that integrates metacognitive processes and key measures of motivation.

processes and sources of motivation that precede efforts to learn and influence students' preparation and willingness to self-regulate their learning. The performance phase involves processes that occur during learning and affect concentration and performance, and the self-reflection phase involves processes that follow learning efforts but influence a learner's reactions to that experience. These self-reflections, in turn, influence forethought regarding subsequent learning efforts, which completes the self-regulatory cycle. It should be noted that the length of a student's self-regulatory cycles will vary based on the frequency and timing of feedback, which in turn depends on outside sources, such as receiving a quiz grade, as well as personal sources, such as keeping a diary.

Forethought Phase

This self-regulatory phase is composed of two major categories: task analysis processes and sources of self-motivation. Task analysis involves decomposing a learning task and its context into constituent elements, and constructing a personal strategy from prior knowledge of these elements (Winne & Hadwin, 1998), such as breaking a math problem down into sequential steps. Task analysis involves two key parts: setting goals and strategic planning. Goal setting refers to specifying the outcomes that one expects to attain, such as solving a page of decimal problems during a one-hour study session (Locke & Latham, 2002). Strategic planning refers to choosing or constructing advantageous learning methods that are appropriate for the task and environmental setting. When students link their strategic plans for learning to short and long-term goals in a sequential or hierarchical system, they can practice effectively by themselves over long periods of time.

Because forethought is anticipatory, it depends on a number of key sources of self-motivation, such as self-efficacy perceptions, outcome expectations, intrinsic interest, and a learning goal orientation. Each of these key sources of motivation has been linked to goal setting and strategic planning. For example self-efficacy, which is defined as beliefs about one's capabilities to learn or perform at designated levels, has been shown to predict students' goals and strategic choices (Zimmerman, Bandura, & Martinez-Pons, 1992). Self-efficacy beliefs can affect performance phase processes directly, such as one's choice of activities, effort, and persistence. A student's self-efficacy perceptions can affect his or her use of learning strategies in diverse areas, such as writing (Schunk & Swartz, 1993), time management (Britton & Tessor, 1991), resistance to adverse peer pressures (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996), and self-monitoring (Bouffard-Bouchard, Parent, & Larivee, 1991).

A second important source of self-motivation, outcome expectancies, refers to beliefs about the ultimate ends of one's performance, such as receiving social recognition or obtaining a desirable employment position. Students' outcome expectations depend on their knowledge or awareness of various outcomes, such as potential salaries, quality of life, and social benefits of a profession. Although the positive effect of attractive outcomes is well established, these expectations also depend on self-efficacy beliefs. For example, a student may believe that it is beneficial to have business accounting skills. However, lacking a sense of efficacy to learn these skills, he or she would not be motivated to enroll in a college accounting course. Thus, students' motivation to self-regulate their learning may be influenced by multiple motives, such as self-efficacy and outcome expectations.

A third source of students' forethought phase motivation is their task interest or valuing. This motive refers to one's liking or disliking a task because of its inherent properties rather than for its instrumental qualities in gaining other outcomes. Deci and Ryan (1985) refer to this motive as intrinsic motivation whereas Wigfield and Eccles (2002) refer to a comparable motive as interest value. Hidi and Renninger (2006) describe a similar motive

as interest, which they view as a cognitive and affective predisposition to re-engage with the content of a particular class of objects, activities, or ideas. Research also shows that task interest or valuing can influence students' choice of learning strategies and well as their achievement goals (Ainley, Corrigan, & Richardson, 2005). Undoubtedly, students' task analytic processes are associated with task value/interest sources of motivation.

A fourth source of motivation to self-regulate is students' goal orientation, which involves their beliefs or feelings about the purpose of learning. Although research on students' goal orientations has led to variations in the names and number of goal orientations by prominent theorists, there is a consensus regarding the purpose of a learning goal orientation and a performance goal orientation. Students who adhere to a learning goal orientation seek to improve their competence via learning, whereas students who adhere to a performance goal orientation seek to protect their competence perceptions via favorable comparisons with the performance of others. According to Dweck and Leggett (1988), students' learning goal orientation stems from a belief that their mental ability can be modified incrementally, whereas a performance goal orientation stems from a belief that their mental ability is a fixed entity.

Students' goal orientation is an important predictor of forethought phase strategic planning. For example, in one study, college students with a strong learning goal orientation chose and implemented advantageous "deep" learning strategies more frequently than students with a weak learning goal orientation (Grant & Dweck, 2003). Strong learning goal students also displayed superior self-reflection processes to students with a strong performance goal orientation. These students recovered more quickly from poor performance on the first exam in the course and displayed higher performance by the end of that course. Research also shows that mastery goals are significantly correlated with measures of intrinsic motivation (Harackiewicz, Barron, & Elliot, 1998).

It should be noted that the motivational construct of goal orientations differs from metacognitive goal setting in the following way: Goal setting commits a student to a specific academic outcome at a particular point in time, such as completing an essay in two days, whereas goal orientation does not target a specific outcome in time. In terms of its role in self-regulation, goal setting produces an explicit feedback loop that requires self-evaluation on a specific time. By contrast, a goal orientation is an open-ended commitment to engage in learning or performance activities.

Performance Phase

This second phase of self-regulation involves two major categories: self-control and self-observation methods. Students' use of self-control methods involves a variety of task-specific as well as general strategies. Task strategies refer to developing a systematic process for addressing specific components of a task, such as creating steps for multiplying fractions in math or for putting in golf. Included among general self-control strategies are the following: self-instruction, imagery, time management, environmental structuring, help-seeking methods, interest incentives, and self-consequences. Self-instruction refers to overt or covert descriptions of how to proceed as one executes a task, such as self-questioning as one reads textual material. Although the effectiveness of one's verbalizations depends on their quality and execution, research has shown that verbalizations can improve students' learning (Schunk, 1982).

Imagery is a self-control strategy that involves forming mental pictures to assist learning and retention, such as converting textual information into visual tree diagrams, flow charts, and concept webs. These graphical representations enable students to retrieve stored information in non-verbal images. The self-control method of time management

refers to strategies for accomplishing learning tasks on schedule, such as setting specific task goals, estimating time requirements for those tasks, and monitoring progress in attaining those goals. There is evidence that students as young as elementary school age (Stoeger & Ziegler, 2008) and as mature as college age (Schmitz & Wiese, 2006) can profit from training in time management strategies.

Environmental structuring is a self-control method for increasing the effectiveness of one's immediate environments, such as using a computer to write an essay because the word processing program provides feedback regarding spelling and grammar. Help-seeking is a method of self-control that involves soliciting assistance when learning or performing, such as seeking out a tennis coach to assist one to master the service motion. To the casual observer, help-seeking appears to be the antithesis of self-control because assistance is sought from others. However, research has shown that poor achievers are very reluctant to seek advice from others, perhaps because they are unsure of what to ask for, when to ask for it, and who to approach (Karabenick, 1998; Newman, 2007). From this perspective, help-seeking can be viewed as a social form of information seeking.

Some methods of self-control are designed to enhance motivation rather than metacognitive processes (Wolters, 2003). For example, the strategy of interest enhancement seeks to render mundane tasks more attractive by increasing game-like qualities, such as competing with a classmate in recalling foreign language words. Another motivational strategy is self-consequences, which involves setting rewarding or punishing contingencies for oneself, such as putting off the task of answering email messages until after completing a homework reading assignment in chemistry (Zimmerman & Martinez-Pons, 1986). Although this list of self-control methods is illustrative rather than exhaustive, it conveys the range of self-regulatory strategies that have been used to enhance students' learning and performance of academic and non-academic skills.

Whether specific or general, all of these strategies need to be adapted based on learners' outcomes, and this is why self-observation plays such a central role in students' efforts to self-control their performance. There are two key forms of self-observation: metacognitive monitoring and self-recording. Metacognitive monitoring or self-monitoring refers to informal mental tracking of one's performance processes and outcomes, such as one's learning processes and their effectiveness in producing learning. By contrast, self-recording refers to creating formal records of learning processes or outcomes, such as a graph of a student's spelling errors in his or her written assignments. Creating a record of one's efforts to learn is advantageous for learners because it increases the reliability, specificity, and time span of self-observations. Self-recording can also include information regarding the conditions that surround a personal event, such as when time management records include descriptions of distracting phone calls from a friend.

Tracking one's performance is difficult when the amount of information involved in complex academic performance exceeds a person's mental capacity. An overload can produce disorganized or cursory metacognitive monitoring but can be avoided if students learn to selectively track key processes, such as the quality of their notes in a history class. Self-monitoring can be improved by recording because these records decrease students' reliance on recall and enable them to discern and interpret subtle changes in performance over time (Zimmerman & Kitsantas, 1999).

Self-Reflection Phase

This phase is composed of two categories of response: self-judgments and self-reactions. A key form of self-judgment is self-evaluation, which refers to comparisons of one's performance with a standard. Bandura (1986) has identified three major types of evaluative

standards: prior levels of performance, mastery of all components of a skill, and social comparisons with the performance of others, such as siblings, peers, and classmates. According to this analysis, learners often have some discretion to select standards to evaluate their performance, such as when a failing student sets a “C” as a personal grade goal. Setting a higher but unrealistic goal (e.g., an “A”) will ultimately undermine that student’s motivation to continue striving if the feedback is negative. Students’ goal setting during the forethought phase will influence the standard that they use to self-evaluate during the self-reflection phase. It is important to note that a student’s choice of a particular standard can greatly affect their perceived outcomes and subsequent motivation. For example, a student’s use of prior performance levels as a standard of comparison has the advantage of committing him or her to self-improvement rather than to surpassing the performance of other students who may have started with an advantageous level of skill.

A second form of self-judgment that plays an important role in understanding cycles of self-regulation are causal attributions. These attributions are defined as beliefs about the causal implications of personal outcomes, such as one’s fixed ability, effort, and use of strategies. A number of researchers (Schunk, 2007; Weiner, 1992) have expressed concern that certain types of attributions for performance can easily undermine self-motivation. For example, attributing errors to uncontrollable factors, such as a lack of talent or ability, prompts learners to react negatively to setbacks and discourages efforts to improve. However, attributing errors to controllable factors, such as use of a particular strategy, can sustain motivation during periods of subpar performance (e.g., Zimmerman & Kitsantas, 1997, 1999). The three-phase social cognitive model of self-regulation places causal attributions and evaluations together as self-judgments because of their conceptual interdependence. For example, if a student evaluates his or her academic performance as successful, then almost any causal attributions for success will increase motivation for subsequent efforts to learn. By contrast, if a student evaluates his or her performance as unsuccessful, then attributions to uncontrollable causes, such as fixed ability, will diminish his or her motivation. Thus, the motivational impact of an attribution depends directly on a student’s evaluative judgments.

A second key category of the self-reflection phase is composed of two forms of self-reaction: self-satisfaction and adaptive/defensive decisions. Self-satisfaction is defined as cognitive and affective reactions to one’s self-judgments. Self-satisfaction has been widely studied because students prefer learning activities that previously led to satisfaction and positive affect, and they tend to avoid those that produce dissatisfaction and negative affect, such as anxiety (Bandura, 1991). Adaptive decisions refer to students’ willingness to engage in further cycles of learning by continuing their use of a strategy or by modifying it. By contrast, defensive decisions avoid further efforts to learn in order to shield a student from future dissatisfaction and aversive affect. Helplessness, procrastination, task avoidance, cognitive disengagement, and apathy are all forms of defensiveness. It is important to note that both forms of students’ self-reactions are dependent on self-judgments during the self-reflection phase. For example, favorable self-evaluations of one’s performance and attributions to controllable causes will in turn lead to increased self-satisfaction and continued efforts to learn adaptively.

These self-reactions affect forethought phase processes in a cyclical fashion during further efforts to acquire proficiency. For example, positive self-satisfaction reactions are expected to enhance positive motivational beliefs about further efforts to learn, such as a more positive perception of self-efficacy about eventually mastering the academic skill, a stronger learning goal orientation (Dweck & Leggett, 1988), and greater intrinsic interest in the task (Zimmerman & Kitsantas, 1997). Because of their cyclical interdependence, self-regulatory processes can become self-sustaining in that processes and beliefs in each

phase create inertia that can facilitate or undermine efforts to learn during subsequent phases. For example, adaptive or defensive decisions can affect students' goals and strategic planning during subsequent cycles of learning. Thus, these altered self-motivation beliefs, goals, and planning form the basis for further self-regulatory efforts.

One of the purposes of the three-phase model of self-regulation was to identify specific metacognitive processes and motivational sources and to explain their cyclical interrelation during ongoing efforts to learn in real contexts. An unexpected by-product of efforts to test this model is the realization that the most prominent sources of motivation involve personal beliefs and feelings that precede, accompany, or follow efforts to learn. For example, self-efficacy refers to a student's a priori confidence about acquiring a skill, whereas task valuing/interest concerns a student's liking of the learning task. A learning goal orientation addresses the purpose of learning, and attributions refer to post hoc judgments of the causes of a student's learning. Self-satisfaction and adaptation deal with students' self-reactions to their ongoing attempts to learn. The point is that any complete accounting of a student's efforts to self-regulate should include not only metacognitive processes but also his or her motivational beliefs and feelings about learning at various points during cyclical feedback loops. Thus, these motivational beliefs are both a cause and an effect of a student's efforts to learn metacognitively.

Microanalytic Measures and Cyclical Analyses to Assess SRL

Zimmerman and his colleagues developed a methodology for assessing most metacognitive and motivational processes of SRL during ongoing efforts to learn. This microanalytic approach involves assessing an individual student's responses to questions at key points before, during, and after learning. Students' answers to these open- or closed-ended questions yield both qualitative and quantitative data respectively. The questions are brief and task-specific in order to minimize disruptions in learning, and they are contextually-specific, which can increase their validity. For example, there is research indicating that microanalytic measures of self-regulation are highly predictive of performance differences between expert, non-expert, and novice athletes (Cleary & Zimmerman, 2001; Kitsantas & Zimmerman, 2002). Novices in these studies displayed deficiencies in not only the quantity and quality of self-regulatory processes but also their motivational beliefs. This microanalytic methodology is classified as an *event* measure of SRL, which is defined as a temporal entity that has a beginning and an end (Zimmerman, 2008). Because a self-regulation event, such as using a strategy, occurs in sequence to events that precede and follow it, these measures are sensitive to change and can capture correlational as well as causal relations between metacognitive processes and motivational feelings and beliefs.

For example, microanalytic measures have revealed evidence of high positive correlations among within-phase measures of motivation, such as between attributions to strategy use and self-satisfaction reactions in research on writing revision. In a study of these issues, Zimmerman and Kitsantas (1999) reported evidence of sequential causality in students' forethought phase strategic planning, performance phase learning, and self-reflection phase attributions and feelings of satisfaction with their writing. Finally, it was found that self-reflection phase measures of motivation were predictive of forethought phase motivational beliefs, such as self-efficacy and task interest regarding additional cycles of learning. Clearly, students' use of high quality SRL processes can lead to enhanced motivation to continue additional cycles of learning.

Microanalytic measures of SRL processes and sources motivation have also been used to investigate students' learning of athletic skills, such as free throw shooting, volleyball serving, and dart throwing. These measures of metacognition and motivation revealed

significant differences among experts, non-experts, and novices (Cleary & Zimmerman, 2001; Kitsantas & Zimmerman, 2002). When experts were compared to non-experts and novices, they reported the greatest use of metacognitive processes and the most positive motivational beliefs. In the study of volleyball serving, microanalytic measures of self-regulation during a practice phase were combined and used to predict accuracy in serving during a post-test (Kitsantas & Zimmerman, 2002). The combined measures predicted over 90% of the variance in post-test skill. Clearly, microanalytic measures of self-regulatory processes and motivational beliefs during students' efforts to learn have disclosed some impressive findings! Although high levels of expertise take years to develop (Ericsson, 2006), there is recent evidence (Cleary, Zimmerman, & Keating, 2006) that novices who were taught multi-phase SRL strategies for basketball free throw shooting displayed significantly greater athletic skill and improved motivational beliefs during relatively brief practice sessions than novices in an untutored control group.

Becoming a Proactive Self-Regulator

Microanalytic research of Zimmerman and colleagues has revealed that even novices attempt to self-regulate their learning in some way, so it can be asked: How should researchers describe self-enhancing and self-defeating cycles of learning according to a three-phase model of self-regulated learning? We define students who focus on and engage in productive forethought before they attempt to learn as proactive self-regulators. Alternatively, we define students who rely primarily on self-reflections based on performance phase outcomes as reactive self-regulators. Unfortunately, the latter students handicap themselves because of their failure to analyze the task, to set specific goals, and plan an effective strategy before attempting to learn (Bandura & Schunk, 1981). By contrast, proactive self-regulators set specific learning goals that are linked closely to strategic planning. There is growing evidence that learners who focus initially on learning processes learn much more effectively than learners who focus initially on learning outcomes (Schunk & Schwartz, 1993; Zimmerman & Kitsantas, 1996).

In addition to this empirical evidence demonstrating the advantage of process goals versus outcome goals, there is extensive anecdotal evidence that learners who set process goals learn more effectively. For example, Paul Annacone (2008), a veteran tennis coach who worked with a highly ranked British tennis player, Tim Henman, described not only the metacognitive advantages of setting process goals but also the motivational benefits, such as attributions to strategy use and self-satisfaction.

Think about tactics, not results. As I found out early on with Tim, the score can get in the way of your tennis. Tim and I used to talk a lot about being process-oriented rather than results-oriented. If you go into a match ready to fight and win points, you might forget to think about what it is you should do to win those points. Tim really latched onto thinking more about tactics than score. He learned to put patterns together in his mind of how he wanted to play points, rather than focusing on whether he won them or not. Perhaps the biggest benefit to learning to think like Tim: When you've committed to a strategy and tried your best to execute it, you'll have no regrets.

(Annacone, 2008)

Does Annacone's advice mean that setting outcome goals are inherently counter-productive when students seek to learn on their own? An answer to this question requires a detailed analysis of metacognitive as well as motivational processes. We have already discussed research showing that outcome goals that are specific, proximal, and challenging

are more effective than outcome goals that are unfocused, distal, and easy (Bandura & Schunk, 1981; Locke & Latham, 2002). Clearly, both high-quality outcome goals and process goals are advantageous, but how should they be structured to optimize learning?

Zimmerman and Kitsantas (1997) hypothesized that with complex tasks, process goals are advantageous during initial learning because of their close linkage to strategic planning and implementation. In terms of their impact on forethought, process goals are designed to incorporate strategic planning—combining two key task analysis processes. With studying and/or practice, students will eventually use the strategy automatically. Automization occurs when a strategy can be executed without close metacognitive monitoring. At the point of automization, students can benefit from outcome feedback because it helps them to adapt their performance based on their own personal capabilities, such as when a basketball free throw shooter adjusts their throwing strategy based on their height. However, even experts will encounter subsequent difficulties after a strategy becomes automatic, and this will require them to shift their monitoring back from outcomes to processes. This is one of the reasons that professional tennis players hire coaches, such as Paul Annacone.

To validate this metacognitive account and establish its motivational power, Zimmerman and Kitsantas (1997, 1999) conducted two experimental studies of process and outcome goals, one with an athletic task and another with an academic task, namely writing. In a study of writing revision with adolescent girls, all participants were initially taught a three-step revision strategy through observation and emulation of a model. Following training, a practice session was conducted. Girls in the process goal group focused on strategic steps for revising each writing task. By contrast, girls in the outcome goal focused on decreasing the number of words in their revised passages. The presence of unnecessary words reduced post-test writing revision scores. Girls in a shifting goal group started with process goals and changed to outcome goals after automatization occurred. Half of the girls in each goal group were asked to self-record their processes or outcomes, and this was expected to enhance both types of performance phase learning.

It was found that girls who shifted goals from processes to outcomes after having achieved automatization surpassed the writing revision skill of girls who adhered exclusively to process or outcome goals. Girls who focused on outcomes exclusively displayed the least writing skill, and self-recording enhanced writing acquisition for all goal setting groups. In addition to their acquisition of superior writing skill, girls who shifted their goals displayed advantageous forms of self-motivation, such as greater attribution to controllable causes (i.e., strategy use), enhanced self-satisfaction, more optimistic self-efficacy beliefs, and greater task interest. The motivational advantages of shifting one's goals metacognitively at the point of automatization were replicated in research on acquisition of an athletic skill (Zimmerman & Kitsantas, 1997).

Self-Efficacy and Self-Evaluative Calibration

The close linkage between metacognitive processes and sources of motivation is especially evident in recent research on self-efficacy calibration. As was discussed earlier, there is extensive research documenting significant positive correlations between the *strength* of students' self-efficacy beliefs and their academic achievement (Pajares, 1996; Zimmerman, 1995). According to a strength hypothesis, students' reports of higher self-efficacy should lead to higher motivation and achievement. For example, Schunk (1984) found that students' self-efficacy beliefs predicted their persistence during learning as well as their success in acquiring math skills.

However, there is recent research that focuses on another dimension of self-efficacy—the

calibration of students' self-efficacy beliefs. Calibration is a measure of metacognitive monitoring based on the disparity between one's sense of efficacy about performing a particular task and one's actual performance. This calibration issue has arisen in cases where test item-specific measures of self-efficacy have been employed, such as a student's answer to a math problem. There is evidence that students often overestimate their efficacy judgments—and in some cases quite substantially (Klassen, 2002; Pajares & Miller, 1994). Underestimates of self-efficacy are much less frequent, but their effects on students' motivation can be particularly disabling because they can undermine students' effort, persistence, and choice of challenging academic tasks (Bandura, 1986; 1997). Ideally, students' self-efficacy judgments should not greatly exceed their actual capability because over-optimism can lead to insufficient efforts to learn (Ghatala, Levin, & Pressley, 1989). Well-calibrated people are accurate in judging their capability to perform a task learn more effectively (Schunk & Pajares, 2004).

Although research on self-efficacy calibration is quite recent, there is an older literature on self-evaluation calibration, which involves students' judgments of knowing *after* completing a test item. Research on self-evaluation has also reported extensive evidence that students' accuracy in judging the correctness of their answers is only slightly above chance. However, better accuracy is related to better performance (Chen, 2003). Researchers have found that low-achieving students are less accurate and have a greater tendency toward overconfidence than high-achieving students who tend to be slightly underconfident (Bol & Hacker, 2001). Intervention studies designed to improve students' metacognitive accuracy have been largely unsuccessful, implying that these self-evaluative judgments may be resistant to change (Bol & Hacker, 2001; Hacker & Bol, 2004).

Among possible reasons for the resistance of overconfident students to change, motivational concerns figure prominently. It has been speculated that low-achieving students may resist calibration training because of a self-serving attributional style. That is, optimistic misjudgments of learning are blamed on others whereas accurate self-evaluative judgments are attributed to themselves (Hacker & Bol, 2004). However, inaccurate perceptions of one's capability can have negative consequences. Students who grossly overestimate their capabilities may attempt to solve difficult problems and will experience failure, which can decrease their motivation for engaging in further learning (Bandura, 1986; Schunk & Pajares, 2004). Pajares (1996) has discouraged interventions designed to lower students' self-efficacy judgments, but he has advocated improving the accuracy of their efficacy judgments by helping them metacognitively to comprehend what they know and do not know. Stone (2000) and Zimmerman (1990) hypothesized that self-regulated learners are well calibrated. Students who engage in high-quality task analysis, choose an effective strategy, and implement it as a process goal are unlikely to greatly overestimate their self-efficacy.

An Intervention Study to Improve Students' Calibration and Self-Reflection

Zimmerman and his colleagues (Zimmerman, Moylan, Hudesman, White, & Flugman, 2008) sought to compare the effects of self-reflection training on students' metacognitive calibration judgments and motivational beliefs in an intervention study with at-risk technical college students who are studying math. This recent study was designed to help these struggling learners interpret their academic grades as sources of self-reflective feedback rather than as indices of personal limitation. College students in developmental (remedial) math or introductory college-level math courses were randomly assigned to either an experimental or a control classroom of their respective courses. In self-reflection

classrooms, teachers used modeling techniques and assessment practices designed to enhance self-reflection processes, while students were given frequent, ongoing opportunities to improve their math achievement using a self-reflective feedback form to self-regulate their math learning and problem solving.

Self-reflective feedback was provided every two to three class sessions. Students in the intervention classrooms were administered a 15–20 minute quiz involving four to five math problems as a vehicle for frequent feedback to students and teachers. Quizzes required students to make task-specific self-efficacy judgments before solving individual problems and self-evaluative judgments after attempting to solve each math problem. After receiving graded quizzes from the instructor, students in the intervention group had the opportunity to correct quiz errors by completing self-reflection forms and to receive grade point incentives.

The self-reflection form was designed to guide students' self-reflection processes regarding erroneous answers to items on a mathematics quiz and cyclical self-regulatory efforts to solve transfer problems. For example, the self-reflection form required students to compare their self-efficacy and self-evaluative judgments with their outcome on the quiz item, explain their ineffectual strategies, develop a new more effective strategy, and indicate their confidence for solving a new problem. If the student failed to solve the problem, they were encouraged to seek assistance from an instructor, a tutor, or a peer. To help students understand and complete the self-reflection forms, instructors initially modeled the completion of the forms, and they supported their students with in-class group and individual practice. The rubric for scoring the quality of the students' answers on the self-reflection forms was given to the students by the instructors.

As expected, Zimmerman and his colleagues found substantial evidence of overconfidence by these at-risk technical college students, despite low levels of math achievement. The attrition and failure rate in math classes at this public technological college has been very high. The process of helping students self-reflect on their errors in mathematical learning and problem solving was designed to shift perceptions of assessment feedback from being an end-point to instead being viewed as a source for enhanced self-regulation.

The results revealed that students receiving self-reflection training outperformed control group students on instructor-developed tests and were better calibrated in their task-specific self-efficacy beliefs before solving problems and in their self-evaluative judgments after solving problems. Students in the conventional instruction classes tended to overestimate their perceived competence in mathematics. As predicted from a social cognitive model of self-regulated learning, evidence was found of significant relations among self-regulated learning processes. Specifically, students' self-efficacy and self-evaluation judgments regarding performance on periodic tests were positively correlated with their math test achievement. Additionally, students' self-efficacy for final exam achievement, their standards for self-satisfaction, and their self-reported learning strategy use were each positively correlated with final exam achievement. Furthermore, among students receiving self-regulated learning training, higher amounts of error correction were predictive of math achievement and metacognitive calibration.

These findings suggest that a self-regulated learning intervention designed to improve students' self-reflection did improve the accuracy of students' self-monitoring of their mathematical problem-solving performance. This, in turn, had positive effects on their achievement in entrance-level courses in college mathematics. Receiving an academic grade is an important transactional event between a teacher and a student. From a student's standpoint, feedback from traditional forms of math assessment is usually problematic regarding self-regulation and is often perceived as punishing. An instructional and assessment approach that emphasizes interpreting errors from a self-regulatory perspective

is more likely to empower students to respond adaptively to academic feedback. Although this intervention demonstrated the effectiveness of closely linked metacognitive and motivational components, it did not address the issue of their separate and combined effectiveness in comparison to other studies. We turn to this issue next.

Evaluating the Academic Effectiveness of Metacognitive and Motivational Training

In a recent meta-analysis, Dignath, Buettner, and Langfeldt (2008) reviewed 48 studies derived from 30 articles on the effectiveness of self-regulatory training with primary school students. The conceptual framework for selecting studies involved motivational as well as cognitive and metacognitive processes. To be included in the meta-analysis, the intervention had to be conducted in a classroom setting by either teachers or researchers. Interventions that lacked the following characteristics were excluded: those that did not include strategy instruction, lasted fewer than two days, lacked a control group, or had fewer than 10 participants. Studies that were situated in computer contexts were also excluded. To insure comparable student comparisons across the studies, authors excluded gifted and learning disabled students. Finally, the review did not include studies published before 1992 because that was the date of the last extensive review of research on self-regulated learning strategies (Hattie, 1992).*

The selected studies were classified based on a variety of issues. These included the types of strategies that were taught, such as cognitive, metacognitive, and motivational. Cognitive strategies referred to direct regulation of learned information, such as a math calculation strategy. Metacognitive strategies referred to second-order cognitions designed to control, monitor, and evaluate learning and cognitive activities. Motivational strategies included self-efficacy, attributional orientation, action control methods, and feedback. The studies were also classified based on the theoretical model that guided the study, including metacognitive, motivational, and social cognitive/constructivist as well as combinations of these theories.

The results revealed that most of these self-regulation intervention studies produced not only gains in students' academic performance but also improvements in their strategic behavior and motivation. The mean effect size of self-regulated learning training was 0.61 for overall academic performance, 0.73 for enhanced strategy use, and 0.76 for improved motivation. It should be noted that for these tests, Cohen (1988) classifies effect sizes of 0.80 and above as large, effect sizes between 0.50 and 0.79 as medium, and effect sizes between 0.49 and below as small.

Regarding the effects related to the theoretical background for the study on students' overall academic performance, interventions based on social-cognitive/constructivist theories had a large effect size (0.95), whereas interventions based on motivation theories exerted the smallest effect size (0.33). Interventions based on metacognitive theories were medium in their effect size (0.58). Interestingly, intervention studies that combined social-cognitive/constructivist and metacognitive theories displayed the largest effect (1.44).

The data were also analyzed for the effects of the type of strategies that were implemented to enhance overall academic performance. Note that the theoretical framework behind interventions did not necessarily correspond with the strategies emphasized in

* In this study, Dignath and Buettner (in press) used regression analyses to assess the relation between moderator variables and effect sizes and did not report the effect sizes for these variables. For this reason, we described these findings without reporting the size of the specific effects.

training. It was found that interventions that relied on cognitive strategies produced only low effects. Although the effect of adopting a purely motivational model of self-regulated learning on overall academic performance was small, interventions that emphasized motivational strategies (1.36) or a combination of metacognitive and motivational strategies (1.23) exerted large effects on students' overall academic attainments.

Regarding the metacognitive reflection strategies that were used to enhance students' overall academic performance, the highest effect sizes were found with interventions that provided students with knowledge about strategies (0.91) and demonstrated the benefits of applying the trained strategies or triggered metacognitive reasoning (0.78).

The data were also analyzed according to the effects of different types of motivational strategy on students' overall academic performance. The largest effect size was produced by a feedback strategy (1.41). In addition to its impact on academic learning, the feedback strategy had a very large impact on the students' motivation (1.40). By contrast, causal attribution strategies (0.64) and an action control strategies (0.48) produced medium-sized effects. Clearly the most effective training methods provided students with feedback about their strategic learning. These results provide support for the use of self-reflection forms like those used by Zimmerman and his colleagues (2008) in their intervention study.

Concerning the role of context factors in these self-regulated learning interventions, there is evidence that the school subject matter is important. The size of self-regulatory training effect on students' math outcomes was large (1.00), but the effect on their reading and writing was small (0.44). Another contextual factor involves the impact of the length of interventions on students' academic learning, but the meta-analysis revealed no evidence that the duration of the interventions is predictive of academic outcomes. A third contextual factor involves the person who implements a self-regulatory intervention. The effect size for researcher-directed interventions was large (0.87) whereas the effect size for teacher-directed interventions was small (0.46). In our experience, teachers adhere less closely to intervention guidelines than researchers. A fourth context factor involves the methods used to assess students' use of self-regulatory processes. The largest effect size was found for questionnaire measures (0.86). Medium effect sizes were found regarding the use of task tests, simulation tasks, think-aloud records, and observation measures. The lowest effect sizes were produced by multiple-choice tests. The latter tests do not require recall and are a less demanding measure of self-regulated learning interventions (Ghatala, Levin, & Pressley, 1989). A fifth contextual factor involves the role of students' age in their response to self-regulation training. The results showed no age differences in the effectiveness of self-regulated learning instruction. This indicates that even children in the primary grades (1–3) can acquire self-regulatory competence through systematic training.

In a subsequent meta-analysis of self-regulation training studies, Dignath and Buettner (2008) expanded their database on primary school students to include 35 studies of secondary students that contained 94 effect sizes (357 total for both school levels). This comparative investigation revealed the effect of self-regulation training on secondary students' overall academic performance to be 0.54. This training effect was smaller than that of primary students (0.61). Regarding students' learning of specific academic subjects, the authors found interesting differences based on students' level of schooling. For secondary students, the effect of self-regulation training on reading/writing performance was large (0.92) but on math performance was small (0.23). For primary students, the pattern was reversed: The effect of training on reading/writing performance was small (0.44) but on math performance was large (0.96).

In terms of the effects of self-regulation training on students' use of strategies, the effect size was medium for primary students (0.72) but large for secondary students (0.88). It appears that the effects of strategy training increased from primary to secondary school

levels. The effect of self-regulation training on students' motivation was medium in size for primary students (0.75) but small in size for secondary students (0.17). However, there were only six intervention studies of secondary students that included motivation measures, and the authors cautioned readers against drawing conclusions from these limited data. Regarding the effectiveness of self-regulation instructors, both primary and secondary students profited more when training was conducted by researchers rather than regular teachers.

In terms of the effect of the theoretical background of self-regulation training, secondary students responded differently to primary students. For secondary students, the effect of metacognitive interventions on overall academic performance was greater than other interventions whereas for primary school students, the effect for social-cognitive/constructivist interventions was greater. The effect of motivational interventions was smaller than the preceding interventions for both primary and secondary students.

Concerning the impact of different forms of self-regulation training on academic performance, the effects of metacognitive and motivational strategies were each larger than other strategies for primary students whereas the effects for metacognitive reflection or motivational strategies were each larger for secondary students. Clearly, motivational strategies play an important role in the academic success of both primary and secondary students.

Conclusions

This chapter addressed the role of motivation in students' efforts to self-regulate more effectively. Historically, interventions involving metacognitive measures of learning fared poorly in inducing strategy maintenance and transfer. A case was made that proactive self-regulation depends on the presence of important sources of motivation. We described how a number of highly regarded measures of motivation could be linked conceptually to key metacognitive processes within a social cognitive model of self-regulated learning. Empirical support for this three-phase cyclical model was discussed, especially from research involving event measures of self-regulation during learning episodes.

The intersection of metacognitive and motivational measures has led to the consideration of a number of emergent issues—both methodological and conceptual. One issue concerned the predictive power of microanalytic assessment of self-regulation during the course of functioning. High levels of correlation were reported in research on academic as well as athletic functioning. A second issue was how proactive learners can be developed. The three-phase model was used to explain links between students' process and outcome goal setting (forethought phase), strategy automatization (performance control phase), and motivational reactions to these goal outcomes (self-reflection phase).

A third issue involved research on students' calibration bias regarding two key self-beliefs: self-efficacy and self-evaluation. Calibration is a measure of metacognitive monitoring that is derived from these two widely used measures of motivation, and there is evidence that measures of calibration bias are associated with poorer learning outcomes. A fourth issue concerned whether students' calibration as well as their academic outcomes could be improved through self-reflection training. A recent investigation was successful in enhancing the calibration and achievement of students who were at high academic risk.

A final issue involved the relative effects of metacognitive and motivational components of self-regulated learning. This issue was investigated in meta-analyses of self-regulated learning with primary and secondary school students. Interventions based on social-cognitive/constructivist theoretical backgrounds had a substantial effect on primary students' academic performance whereas interventions based on metacognitive theoretical

backgrounds had a substantial effect on secondary students' academic performance. Interventions based on motivational theoretical backgrounds were less effective for both primary and secondary students. However, the inclusion of motivational strategies in self-regulated learning interventions enhanced the overall academic performance of both primary and secondary students. Clearly, exploring the intersection of metacognition and motivation has opened new windows to our understanding of how students self-regulate and self-sustain their learning.

References

- Ainley, M., Corrigan, M., & Richardson, N. (2005). Students, tasks and emotions: Identifying the contribution of emotions to students reading of popular culture and popular science texts. *Learning and Instruction, 15*, 433–447.
- Annacone, P. (2008). Learning from Tim Henman. *Tennis*. January–February issue. Retrieved from <http://www.tennis.com/print.aspx?id=109326>
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1991). Self-regulation of motivation through anticipatory and self-reactive mechanisms. In R. A. Dienstbier (Ed.), *Perspectives on Motivation: Nebraska Symposium on Motivation* (Vol. 38, pp. 69–164). Lincoln: University of Nebraska Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development, 67*, 1206–1222.
- Bandura, A. & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41*, 586–598.
- Bol, L. & Hacker, D. (2001). The effect of practice tests on students' calibration and performance. *Journal of Experimental Education, 69*, 133–151.
- Borkowski, J. (1985). Signs of intelligence: Strategy, generalization, and metacognition. In S. R. Yussen (Ed.), *The growth of reflection in children* (pp. 105–144). Orlando, FL: Academic Press.
- Bouffard-Bouchard, T., Patent, S., & Larivee, S. (1991). Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students. *International Journal of Behavior Development, 14*, 153–164.
- Britton, B. K. & Tessor, A. (1991). Effects of time management practices on college grades. *Journal of Educational Psychology, 83*, 405–410.
- Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. Markman (Eds.), *Handbook of child psychology*, Vol. III, *Cognitive development* (pp. 77–166). New York: Wiley.
- Chen, P. P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*, 79–92.
- Cleary, T. & Zimmerman, B. J. (2001). Self-regulation differences during athletic practice by experts, non-experts, and novices. *Journal of Applied Sport Psychology, 13*, 61–82.
- Cleary, T. J., Zimmerman, B. J., & Keating, T. (2006). Training physical education students to self-regulate during basketball free-throw practice. *Research Quarterly for Exercise and Sport, 77*, 251–262.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Deci, E. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
- Dignath, C. & Buettner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*, 231–264. .

- Dignath, C., Buettner, G., & Langfeldt, H. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review*, doi:10.1016/j.edurev.2008.02.003.
- Dweck, C. & Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256–273.
- Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. Hoffman (Eds.), *Handbook of expertise and expert performance* (223–241). New York: Cambridge University Press.
- Ghatala, E. S., Levin, J. R., & Pressley, M. (1989). Improving childrens' regulation of their reading PREP time. *Contemporary Educational Psychology*, 14, 49–66.
- Grant, H. & Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology*, 85, 541–553.
- Hacker, D. & Bol, L. (2004). Metacognitive theory: Considering the social-cognitive influences. In D. M. McInerney & S. Van Etten (Eds.), *Research on sociocultural influences on motivation and learning: Big theories revisited* (pp. 275–297). Vol 4. Greenwich, CT: Information Age Press.
- Harackiewicz, J. M., Barron, K. E., & Elliot, A. J. (1998). Rethinking achievement goals: When are they adaptive for college students and why? *Educational Psychologist*, 33, 1–21.
- Hattie, J. A. (1992). Measuring the effects of schooling. *Australian Journal of Education*, 36, 5–13.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Hidi, S. & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41, 111–127.
- Karabenick, S. A. (Ed.). (1998). *Strategic help seeking: Implications for learning and teaching*. Hillsdale, NJ: Erlbaum.
- Kitsantas, A. & Zimmerman, B. J. (2002). Comparing self-regulatory processes among novice, non-expert, and expert volleyball players: A microanalytic study. *Journal of Applied Sport Psychology*, 14, 91–105.
- Klassen, R. (2002). A question of calibration: A review of the self-efficacy beliefs of students with learning disabilities. *Learning Disability Quarterly*, 25, 88–103.
- Locke, E. A. & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57, 705–717.
- Newman, R. S. (2007). The motivational role of adaptive help seeking in self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 315–338). Mahwah, NJ: Erlbaum.
- Pajares, F. (1996). Self-efficacy beliefs in achievement settings. *Review of Educational Research*, 66, 543–578.
- Pajares, F., Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, 86, 193–203.
- Pressley, M. & McCormick, C. B. (1995). *Advanced educational psychology: For educators, researchers, and policymakers*. New York: Harper/Collins.
- Rabinowitz, M., Freeman, K., & Cohen, S. (1992). Use and maintenance of strategies. The influence of accessibility on knowledge. *Journal of Educational Psychology*, 84, 211–218.
- Scarr, S. & McCartney, K. (1983). How people make their own environment: A theory of genotype-environment effects. *Child Development*, 54, 424–435.
- Schmitz, B. & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology* 31, 64–96.
- Schunk, D. H. (1982). Verbal self-regulation as a facilitator of children's achievement and self-efficacy. *Human learning*, 1, 265–277.
- Schunk, D. H. (1984). Enhancing self-efficacy and achievement through rewards and goals: Motivational and information effects. *Journal of Educational Research*, 78, 29–34.
- Schunk, D. H. (2007). Attributions as motivators of self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 245–266). New York: Lawrence Erlbaum Associates.

- Schunk, D. H. & Pajares, F. (2004). Self-efficacy in education revisited: Empirical and applied evidence. In D. M. McInerney & S. Van Etten (Eds.), *Research on sociocultural influences on motivation and learning: Big theories revisited* (pp. 115–138). Vol. 4. Greenwich, CT: Information Age Press.
- Schunk, D. H. & Schwartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18 (3), 337–354.
- Schunk, D. H. & Zimmerman, B. J. (Eds.). (2007). *Motivation and self-regulated learning: Theory, research, and applications*. Mahwah, NJ: Erlbaum.
- Stoeger, H. & Ziegler, A. (2008). Evaluation of a classroom based training to improve self-regulation in time management tasks during homework activities with fourth graders. *Metacognition and Learning*, 3, 207–230.
- Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, 12, 437–475.
- Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Newbury Park, CA: SAGE Publications.
- Wigfield, A. & Eccles, J. S. (2002). The development of competence beliefs and values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 92–120). San Diego: Academic Press.
- Winne, P. H. & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wolters, C. A. (2003). Regulation of motivation: Evaluating an underemphasized aspect of self-regulated learning. *Educational Psychologist*, 38, 189–205.
- Zimmerman, B. J. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review*, 82, 297–306.
- Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 202–231). New York: Cambridge University Press.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical Background, methodological developments, and future prospects. *American Educational Research Journal*, 45 (1), 166–183.
- Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29, 663–676.
- Zimmerman, B. J. & Kitsantas, A. (1996). Self-regulated learning of a motoric skill: The role of goal setting and self-monitoring. *Journal of Applied Sport Psychology*, 8, 69–84.
- Zimmerman, B. J. & Kitsantas, A. (1997). Developmental phases in self-regulation: Shifting from process goals to outcome goals. *Journal of Educational Psychology*, 89, 29–36.
- Zimmerman, B. J. & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. *Journal of Educational Psychology*, 99, 241–250.
- Zimmerman, B. J. & Martinez-Pons, M. (1986). Development of a structured interview for assessing students' use of self-regulated learning strategies. *American Educational Research Journal*, 23, 614–628.
- Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2008). *Enhancing Self-Reflection and Mathematics Achievement of At-risk Students at an Urban Technical College: A Self-regulated Learning Intervention. A final report submitted to the Institute of Education Sciences*, U.S. Department of Education, through Grant R305H060018 to the City University of New York.

Part VIII

Technology

17 Self-Regulated Learning with Hypermedia

Roger Azevedo and Amy M. Witherspoon

University of Memphis, Department of Psychology, Institute for Intelligent Systems

Learning about conceptually-rich domains with open-ended computer-based learning environments (CBLEs) such as hypermedia involves a complex set of interactions among cognitive, metacognitive, motivational, and affective processes (Azevedo, 2005, 2007, 2008; Graesser, McNamara, & VanLehn, 2005; Jacobson, 2008; Moos & Azevedo, in press; Vollmeyer & Rheinberg, 2006; Zimmerman, 2008). Current research from several fields including cognitive and learning sciences provides evidence that learners of all ages struggle when learning about these conceptually-rich domains with hypermedia. This research indicates that learning about conceptually-rich domains with hypermedia is particularly difficult because it requires students to regulate their learning. Regulating one's learning involves analyzing the learning context, setting and managing meaningful learning goals, determining which learning strategies to use, assessing whether the strategies are effective in meeting the learning goals, evaluating emerging understanding of the topic, and determining whether there are aspects of the learning context which could be used to facilitate learning. During self-regulated learning, students need to deploy several metacognitive processes to determine whether they understand what they are learning, and perhaps modify their plans, goals, strategies, and effort in relation to dynamically changing contextual conditions. In addition, students must also monitor, modify, and adapt to fluctuations in their motivational and affective states, and determine how much social support (if any) may be needed to perform the task. Also, depending on the learning context, instructional goals, perceived task performance, and progress made towards achieving the learning goal(s), they may need to adaptively modify certain aspects of their cognition, metacognition, motivation, and affect.

Despite the ubiquity of hypermedia environments for learning, the majority of the research has been criticized as atheoretical and lacking rigorous empirical evidence (see Azevedo & Jacobson, 2008; Dillon & Jobst, 2005; Dillon & Gabbard, 1998; Jacobson, 2008; Jacobson & Azevedo, 2008; Neiderhauser, 2008; Tergan, 1997a, 1997b). In order to advance the field and our understanding of the complex nature of learning with hypermedia environments, we need theoretically guided, empirical evidence regarding how students regulate their learning with these environments. In this chapter, we provide a synthesis of existing research on learning with hypermedia, illustrate the complexity of self-regulatory processes during hypermedia learning, and present a theoretical account of SRL based on the information processing theory (IPT) of SRL by Winne and colleagues (1998, 2001, 2008). We also provide a synthesis of recent research on using SRL as a framework with which to capture and study self-regulatory processes, and provide theoretically driven and empirically based guidelines for supporting learners' self-regulated learning with hypermedia.

Overview of Research on Learning with Hypermedia

Recent reviews and critiques of the research on learning with hypermedia have indicated that most hypermedia learning studies are atheoretical and suffer from poor experimental designs and flawed analyses (see Azevedo, 2007, 2008; Azevedo & Jacobson, 2008; Dillon & Jobst, 2005; Jacobson, 2008). In addition, Dillon and Gabbard (1998) concluded that hypermedia is generally advantageous for tasks that require rapid searching through multiple documents, and also that increased learner control over information is not beneficial to all learners and may actually be detrimental for low-ability learners. More recently Dillon and Jobst (2005) reviewed the current hypermedia research based on design issues which cover structure and interface issues and individual differences. In this section we briefly highlight some of the core issues.

In terms of design issues, research has examined the differences observed when learners use different kinds of information structures such as linear and hierarchical hypermedia information structures. In general, the reviews indicate that structure affects learning depending on how effectively and how fast a learner can move through a hypermedia environment (e.g., Shapiro, 1999). In addition, hierarchical format may be beneficial for most learners as opposed to linear format (Dillon & Jobst, 2005; Shapiro & Niederhauser, 2004). Others have used advanced organizers to visually present the underlying hypermedia structure to the learner, and have included varying levels of learning control (Shapiro, 1999).

As for individual differences, researchers have focused on learners' prior knowledge of the domain or topic, spatial ability, and cognitive style. In general, learners with high prior knowledge of the domain tend to display different navigational patterns (e.g., Lawless & Kulikowich, 1996) and tend to learn more from hypermedia environments (e.g., Moos & Azevedo, 2008a, 2008b). Spatial ability has also been related to easing learners' disorientation, or feeling of being lost in hyperspace, thus facilitating their ability to navigate by building spatial metaphors (e.g., Hegarty, Narayanan, & Freitas, 2002). Some have discussed learning styles, however we will not discuss this issue because there is no scientific evidence regarding this construct. Other research has also focused on learner characteristics. For example, researchers have examined the role of metacognitive skills (Schwartz, Andersen, Hong, Howard, & McGee, 2004) and ability levels (Liu, 2004) in learners' outcomes from learning with hypermedia environments. In conclusion, individual differences remain a major research issue because they have a tremendous impact on learning with hypermedia.

Despite the recent rise of quality research on learning with hypermedia, we also raise several critical issues related to learning with hypermedia environments which have not yet been addressed by cognitive and educational researchers. For example, there is the question of how (i.e., with what processes) a learner regulates his/her learning with a hypermedia environment. Most of the research has used the product(s) of learning (i.e., pre-test/post-test learning gains) to infer the connection between individual differences (e.g., prior knowledge, reading ability), learner characteristics (e.g., developmental level), cognitive processes (e.g., learning strategies used during learning), and structure of the hypermedia environment or the inclusion (or exclusion) of certain system features. In our research, we have adopted SRL because it allows us to directly investigate how task demands, learner characteristics, cognitive and metacognitive processes, and system structure interact during the cyclical and iterative phases of planning, monitoring, and control while learning with hypermedia environments.

Self-Regulated Learning with Hypermedia

Learning about complex and challenging science topics, such as the human circulatory system and natural ecological processes with multi-representational, non-linear, computer-based learning environments (CBLEs), requires learners to deploy key self-regulatory processes (Azevedo et al., 2004a, 2004b, 2005, 2007, 2008; Biswas, Leelawong, Schwartz, & TAGV, 2005; Graesser et al., 2005; Jacobson, 2008; McNamara & Shapiro, 2005; Neiderhauser, 2008). Recent cognitive research with middle-school, high-school, and college students has identified several key self-regulatory processes associated with learning, understanding, and problem solving with hypermedia-based CBLEs. First, there are planning processes such as activating prior knowledge, setting and coordinating sub-goals that pertain to accessing new information, and defining which problem solution steps to perform for accomplishing a complex task. In addition, there are also several monitoring processes that are deployed during task enactment including monitoring one's understanding of the topic, managing the learning environment and other instructional resources necessary to accomplish the learning goals, and engaging in periodic self-assessment (i.e., checking for the correctness of solution steps while solving problems and using this information to direct one's future learning activities). During task performance a learner must also use several effective learning strategies for accomplishing the task such as coordinating several informational sources (e.g., text, diagram, animations), generating hypotheses, extracting relevant information from the resources, re-reading, making inferences, summarizing, and re-representing the topic based on one's emerging understanding by taking notes and drawing. Lastly, the learner must continuously adjust during learning by handling task difficulties and demands such as monitoring one's progress towards goals, and modifying the amount of time and effort necessary to complete the learning task. As such, we (Witherspoon, Azevedo, & D'Mello, 2008) and other colleagues (e.g., Biswas et al., 2005; Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007) emphasize that understanding the real-time deployment of these processes in the context of learning and problem-solving tasks is key to understanding the nature of adaptivity in SRL among learners of all ages and their influence on learning. Therefore, we propose that an IPT theory of SRL will best accommodate the complex nature of learning with hypermedia environments.

Theoretical Framework: Information-Processing Theory of SRL

SRL involves actively constructing an understanding of a topic/domain by using strategies and goals, regulating and monitoring certain aspects of cognition, behavior, and motivation, and modifying behavior to achieve a desired goal (see Boekaerts, Pintrich, & Zeidner, 2000; Pintrich, 2000; Zimmerman & Schunk, 2001). Though this definition of SRL is commonly used, the field of SRL consists of various theoretical perspectives that make different assumptions and focus on different constructs, processes, and phases (Dunlosky & Lipko, 2007; Metcalfe & Dunlosky, 2008; Pintrich, 2000; Winne & Hadwin, 2008; Zimmerman, 2008). We further specify SRL as a concept superordinate to metacognition that incorporates both metacognitive monitoring (i.e., knowledge of cognition or metacognitive knowledge) and metacognitive control (involving the skills associated with the regulation of metacognition), as well as processes related to planning for future activities within a learning episode and manipulating contextual conditions as necessary. SRL is based on the assumption that learners exercise agency by consciously monitoring and intervening in their learning. While most of our research has focused on the cognitive and metacognitive aspects of SRL with hypermedia, we are currently considering the

incorporation of other key processes such as motivation and affect (Moos & Azevedo, 2008a, 2008b, in press).

Recent studies on SRL with open-ended learning environments such as hypermedia (e.g., see Azevedo, 2005, 2008, for recent reviews) have drawn on Winne and colleagues' (Butler & Winne, 1995; Winne, 2001; Winne & Hadwin, 1998, 2008) Information Processing Theory (IPT) of SRL. This IPT theory suggests a four-phase model of self-regulated learning. The goal of this section is to explicate the basics of the model so as to emphasize the linear, recursive, and adaptive nature of self-regulated learning and make the link to its implication for use in studying SRL with hypermedia (see Greene & Azevedo, 2007, for a recent review).

Winne and Hadwin (1998, 2008) propose that learning occurs in four basic phases: (1) task definition, (2) goal-setting and planning, (3) studying tactics, and (4) adaptations to metacognition. Winne and Hadwin's SRL model differs from the majority of other SRL models, in that they hypothesize that information processing occurs within each phase. Using the acronym COPES, they describe each of the four phases in terms of the interactions between a learner's conditions, operations, products, evaluations, and standards. All of the terms except operations are kinds of information used or generated during learning. It is within this cognitive architecture, comprised of COPES, that the work of each phase is completed. Thus, their model complements other SRL models by introducing a more complex description of the processes underlying each phase. It should be noted that Winne and Hadwin's model is similar to other models which focus on the underlying cognitive and metacognitive processes, accuracy of metacognitive judgments, and control processes used to achieve particular learning goals (e.g., see Dunlosky, Hertzog, Kennedy, & Thiede, 2005; Koriat & Goldsmith, 1996; Nelson & Narens, 1990; Thiede & Dunlosky, 1999).

Cognitive and task conditions are the resources available to the person and the constraints inherent to the task or environment. Cognitive conditions include beliefs, dispositions and styles, motivation, domain knowledge, knowledge of the current task, and knowledge of study tactics and strategies. Task conditions are external to the person, and include resources, instructional cues, time, and the local context. Thus, in Winne and Hadwin's model, motivation and context are subsumed in conditions. Conditions influence both standards and the actual operations a person performs.

Standards are multi-faceted criteria that the learner believes are the optimal end state of whatever phase is currently running, and they include both metrics and beliefs. For example, in the task definition phase, a learner might examine a list of teacher-set learning goals for a hypermedia learning task and develop task standards including what needs to be learned (metrics), as well as beliefs about the act of studying itself, such as the depth of understanding required, or how difficult the task will be. Winne and Hadwin use a bar graph to illustrate how a learner actively determines criteria for "success" in terms of each aspect of the learning task, with each bar representing a different standard with varying qualities or degrees. The overall profile of these phase one standards constitutes the learner's goal. These standards or goals are used to determine the success of any operations the learner might perform within each phase.

Operations are the information manipulation processes that occur during learning, including searching, monitoring, assembling, rehearsing, and translating; or as Winne (2001) refers to them, SMART processes. These SMART processes are cognitive in nature, not metacognitive; as such they only result in cognitive products, or information for each phase. For example, the product of phase one is a definition of the task, whereas the product of phase three might be the ability to recall a specific piece of information for a test. These products are then compared to the standards by way of monitoring.

Through monitoring, a learner compares products with standards to determine if phase objectives have been met, or if further work remains to be done. These comparisons are called cognitive evaluations, and a poor fit between products and standards may lead a person to enact control over the learning operations to refine the product, revise the conditions and standards, or both. This is the object-level focus of monitoring. However, this monitoring also has a meta-level information, or metacognitive focus. A learner may believe that a particular learning task is easy, and thus translate this belief into a standard in phase two. However, in iterating through phase three, perhaps the learning product is consistently evaluated as unacceptable in terms of object-level standards. This may initiate metacognitive monitoring that determines that this meta-level information, in this case regarding the actual difficulty of this task, does not match the previously set standard that the task is easy. At this point, a metacognitive control strategy might be initiated to modify (or to update) that particular standard (e.g., “this task is hard”) which might, in turn, affect other standards created during phase two, goal setting. These changes to goals from phase two may include a review of past material or the learning of a new study strategy. Thus, the model is a “recursive, weakly sequenced system” (Winne & Hadwin, 1998, p. 281) where the monitoring of products and standards within one phase can lead to updates of products from previous phases. The inclusion of monitoring and control in the cognitive architecture allows these processes to influence each phase of self-regulated learning.

Overall, while there is no typical cycle, most learning involves recycling through the cognitive architecture until a clear definition of the task has been created (phase one), followed by the production of learning goals and the best plan to meet them (phase two), which leads to the enacting of strategies to begin learning (phase three). The products of learning, for example an understanding of mitosis, are compared against standards including the overall accuracy of the product, the learner’s beliefs about what needs to be learned, and other factors like efficacy and time restraints. If the product does not adequately fit the standard, then further learning operations are initiated, perhaps with changes to conditions such as setting aside more time for studying. Finally, after the main process of learning has occurred, learners may decide to further alter beliefs, motivation, and strategies that make up SRL (phase four). These changes can include the addition or deletion of conditions or operations, as well as minor (tuning) and major (restructuring) changes to the ways conditions cue operations (Winne, 2001). The output, or performance, is the result of recursive processes that cascade back and forth, altering conditions, standards, operations, and products as needed.

Lastly, Winne (2005) states that certain hypotheses can be postulated when adopting a model of SRL. First, before committing to a goal, a learner must recognize the features of the learning environment that affect the odds of success. Second, if such features are recognized, then they need to be interpreted, a choice must be made (e.g., set a goal), and the learner needs to select among a set of learning strategies that may lead to successful learning. If these first conditions are satisfied, the learner must have the capability to apply these learning strategies. If these three conditions are met, then the learner must be motivated to put forth the effort entailed in applying learning strategies. In sum, this model provides a macro-level framework and elegantly accounts for the linear, recursive, and adaptive nature of self-regulated learning with hypermedia. As such, we have over the last few years adapted this model to account for learning about complex and challenging science topics with hypermedia (see Azevedo, 2002, 2005, 2008).

Previous Research on Self-Regulated Learning with Hypermedia: The Role of Monitoring Processes

In our research we have begun to address the theoretical, methodological, conceptual, and educational issues raised by several SRL researchers (e.g., Ainley & Patrick, 2006; Alexander, 1995; Boekaerts & Cascallar, 2006; Efklides, 2006; Pintrich, 2000; Veenman, 2007; Winne & Perry, 2000; Zeidner, Boekaerts, & Pintrich, 2001; Zimmerman, 2008). First, we chose a mixed methodology, using true- and quasi-experimental designs combined with concurrent think-alouds protocols (based on Ericsson, 2006; Ericsson & Simon, 1993) to produce both outcome measures (i.e., shifts in mental models from pre-test to post-test) and process data (i.e., concurrent think-aloud protocols detailing the dynamics of SRL processes during learning). Our primary purpose for using concurrent think-aloud protocols was to map out how SRL variables influence shifts in mental models during learning with a hypermedia environment. Second, we triangulated different data sources, both product (e.g., essay responses, diagram labeling, and outlines of the path of blood on pre-tests and post-tests; matching concepts to definitions) and process data (e.g., think-aloud protocols during learning and video coding to capture non-verbal processes and behaviors), to begin to understand the role of SRL in learning complex scientific topics with hypermedia.*

In addition to the theoretical issues presented previously, we also addressed several methodological issues raised by SRL researchers (Hadwin, Winne, & Stockley, 2001; Pintrich, 2000; Winne, 2001; Winne & Perry, 2000; Zeidner et al., 2000; Zimmerman & Schunk, 2001). Little research has been conducted on the inter-relatedness and dynamics of SRL processes—cognitive, motivational/affective, behavioral, and contextual—during the cyclical and iterative phases of planning, monitoring, control, and reflection during learning with hypermedia environments. However, as Hadwin and colleagues (2001) point out, “[if the] hallmark of SRL is adaptation, then data that consist only of self-report questionnaire data and scales that aggregate responses independently of time and context may weakly reflect, and may even distort, what SRL is” (p. 486). One of the main methodological issues related to SRL that we address in our studies is how learners regulate their learning during a knowledge construction activity. Most of the research focuses on SRL as an aptitude—based on learners’ self-reports of perceived strategy use, resource use, and goal selection that depends on the instructional context (e.g., reading for learning, preparing for a midterm exam). These are mostly correlational studies which assess two or three different goal orientations at the same time, or experimental studies which focus almost exclusively on declarative measures of learning. They assume that differences between pre-test and post-test measures reflect learners’ use of cognitive, motivational and contextual SRL variables (Pintrich, 2000). We therefore extended our methodologies and thereby contributed to an emerging set of trace methodologies, which are needed to capture the dynamic and adaptive nature of SRL during learning of complex science topics and with complex, dynamic hypermedia learning environments.

Monitoring Processes During Learning with Hypermedia

In this section, we present the monitoring processes we have identified in our studies on SRL with hypermedia. Although many of these processes are likely context-independent,

* Concurrent think-aloud protocols have been used extensively by researchers in several areas including expertise, problem solving, and reading and text comprehension, to examine the cognitive and metacognitive processes used during learning and performance (see Ericsson, 2006, for a recent review).

applicable to learning with various types of learning resources, some are most appropriately applied to learning with hypermedia, in situations where learners have control over which content, in which modality, they access at any given moment. However, even though we will present real examples of the use of each of these monitoring processes in the context of the circulatory system, each are applicable to learning in any domain.

As previously mentioned, Winne and colleagues' model provides a macro-level framework for the cyclical and iterative phases of SRL. The data presented in this section provides the micro-level details that can interface Winne's model. In Table 17.1, we present the eight metacognitive monitoring processes we have identified as essential to promoting students' self-regulated learning with hypermedia. Some of these monitoring processes are coded along with a valence, positive (+) or negative (-), indicating the learners' evaluation of the content, their understanding, their progress, or familiarity with the material. For example, a learner might state that the current content is either appropriate (content evaluation +) or inappropriate (content evaluation -), given their learning goals and, according to which valence is associated with the evaluation, make choices about persisting with the current material or seeking other content.*

The first monitoring process we have identified in our research is Feeling of Knowing (FOK). FOK is when the learner is aware of having (+) or having not (-) read or seen something in the past and having (+) or not having (-) some familiarity with the material. An example of a learner using FOK is (excerpt from a participant's transcript): "I learned about this, when did I learn about this, that looks familiar."

Closely related to FOK is the monitoring process Judgment of Learning (JOL). JOL is when a learner becomes aware that he/she does (+) or does not (-) know or understand something he reads. An example of a learner using JOL is: "I don't really understand how that works."

The next monitoring process is Monitoring Use of Strategies (MUS). In MUS, the learner acknowledges that a particular learning strategy he/she has employed was either useful (+) or not useful (-). An example of a learner monitoring use of strategies is: "Yeah, drawing it really helps me understand how blood flows throughout the heart."

Self-test (ST) is when a learner poses a question to him/herself to assess the understanding of the content and determine whether to proceed to re-adjust. An example of a learner self-testing is: "OK, well how does it go from the atrium to the ventricle?"

In Monitoring Progress toward Goals (MPTG), learners assess whether previously set goals have been met (+) or not met (-), given time constraints. An example of a learner MPTG is: "Let's see blood, heart, I've done blood and heart and I've done circulatory." Time Monitoring (TM) involves the learner becoming aware of remaining time which was allotted for the learning task. An example of a learner monitoring his time is: "I still have plenty of time."

Content evaluation (CE) is when the learner monitors the appropriateness (+) or inappropriateness (-) of the current learning content, given their pre-existing overall learning goal and subgoals. An example of a learner evaluating the content is: "This section with the diagram of the heart with all the labels is important for me to understand the different components of the heart." The last monitoring process is Evaluation

* The use of concurrent think-alouds allows researchers to capture and classify particular processes and determine when they are deployed during learning. The classification can be accomplished at different levels of granularity: (a) macro-level (e.g., monitoring process), and micro-level (e.g., JOL) with associated valence (either + or -). However, it should be noted that other classification systems and associated analytical approaches and statistical analyses yield different indices of metacognitive monitoring such as absolute accuracy, relative accuracy, bias, scatter, and discrimination (e.g., see Benjamin & Diaz, 2008; Greene & Azevedo, 2009; Schraw, this volume; Van Overschelde, 2008).

Table 17.1 Monitoring Processes Used by Learners during Self-Regulated Learning with Hypermedia

<i>Monitoring process</i>	<i>Description</i>	<i>Role of monitoring process</i>	<i>Aspects of learning situation monitored</i>
Feeling of Knowing (FOK)	Learner is aware of having/having not read something in the past and having/not having some understanding of it	Monitoring correspondence between learner's pre-existing domain knowledge and the learning resources	*Domain knowledge (cognitive condition) *Learning resources (task condition)
Judgment of Learning (JOL)	Learner becomes aware that he or she does/doesn't know or understand something he or she reads	Monitoring correspondence between learner's emerging understanding and the learning resources	*Domain knowledge (cognitive condition) *Learning resources (task condition)
Monitoring Use of Strategies (MUS)	Learner becomes aware that a particular learning strategy he or she has employed was either useful or not useful	Monitoring efficacy of learning strategies, given learner's expectations of learning results and actual learning results	*Learning strategies (operations) *Expectation of results (standards) *Domain knowledge (cognitive condition)
Self-Test (ST)	Posing a question to oneself about the content to assess one's understanding and determine whether to proceed to re-adjust	Monitoring learner's emerging understanding	*Domain knowledge (cognitive condition) *Expectations of content (standards)
Monitoring Progress Towards Goals (MPTG)	Assessing whether previously set goal(s) has been met, given time constraints	Monitoring fit between learning results and previously set goals for learning results	*Domain knowledge (cognitive condition) *Expectation of results (standards) *Learning goals (products)
Time Monitoring (TM)	Learner is aware of remaining time allotted to learning task	Monitoring the task condition of time	*Time (task condition) *Learning goals (products)
Content Evaluation (CE)	Learner becomes aware that a particular piece of content is appropriate/not appropriate, given learner's global goal and sub-goals	Monitoring appropriateness of current learning content given learner's existing learning goals, both current and overall learning goals	*Learning resources (task condition) *Learning goals (products)
Expectation of Adequacy of Content (EAC)	Assessing the expected usefulness and/or adequacy of content not yet navigated toward	Monitoring appropriateness of available learning content given learner's existing learning goals	*Learning resources (task condition) *Learning goals (products)

of Adequacy of Content (EAC). EAC is similar to CE, in that learners are monitoring the learning content, given their learning goals, but in this process, learners evaluate learning content they have not yet navigated toward. An example of a learner evaluating the adequacy of content is: "Do they have a picture of the blood flow through the heart?"

Self-Regulation Using Monitoring Processes

In this section, we describe the learner's application of these eight monitoring processes within the context of self-regulation with hypermedia. For each monitoring process, we provide the aspects of the learning environment which are evaluated, as well as illustrate them using examples of task and cognitive conditions which may lead to the various monitoring processes, as well as examples of appropriate control mechanisms which might be deployed following the evaluations. FOK is used when the learner is monitoring the correspondence between his or her own pre-existing domain knowledge and the current content. The learner's domain knowledge and the learning resources are the aspects of the learning situation being monitored when a learner engages in FOK. If a learner recognizes a mismatch between his/her pre-existing domain knowledge and the learning resources, more effort should be expended in order to align the domain knowledge and the learning resources. Following more effortful use of the learning material, a learner is more likely to experience/generate more positive FOKs. However, if a learner experiences familiarity with some piece of material, a good self-regulator will attempt to integrate the new information with existing knowledge by employing Knowledge Elaboration. Often, a learner will erroneously sense a positive FOK toward material, and quickly move on to other material, with several misconceptions still intact.

In contrast to FOK, JOL is used when the learner is monitoring the correspondence between his/her own emerging understanding of the domain and the learning resources. Similar to FOK, when engaging in JOL a learner is monitoring his domain knowledge and the learning resources. If a learner recognizes that the emerging understanding of the material is not congruent with the material (i.e., the learner is confused by the material), more effort should be applied to learn the material. A common strategy employed after a negative JOL is re-reading previously encountered material. In order to capitalize on re-reading, a good self-regulator should pay particular attention to elements in a passage, animation, or illustration that confused the learner. When a learner expresses a positive JOL, he/she might self-test to confirm that the knowledge is as accurate as the evaluation suggests. As with FOK, learners often over-estimate their emerging understanding and progress too quickly to other material.

In MUS, a learner is monitoring the efficacy of recently used learning strategies, given his/her expectations for learning results. MUS encompasses a learner's monitoring of learning strategies, expectations of results, and domain knowledge. By noting the learning strategies used during a learning task and the resulting change in domain knowledge, learners can compare this emergent knowledge with their expectations and make changes to the strategies employed accordingly. For example, many learners will begin a learning episode by taking copious amounts of notes, then realize that the learning outcomes from this strategy are not as high as they would have expected. Good self-regulators will then make alterations to their strategy of note-taking such as employ more efficient methods (making bullet points, outlines, or drawings), or even abandon this strategy for another, more successful strategy (e.g., summarizing). However, if a learner realizes that a particular strategy has been especially helpful to his/her learning, he/she should continue to employ this strategy during the learning session.

Learners self-test (ST) to monitor their emerging understanding of the content and the aspects of the learning situation being monitored are the learner's domain knowledge and the learner's expectations of the content. While tackling difficult material (e.g., complex science topics), learners should occasionally assess their level of understanding of the material by "administering" an ST. If the results of this self-test are positive, the learner can progress to new material, but if the learner recognizes, through this self-test, that their

emergent understanding of the current material is not congruent with what is stated in the material, they should revisit the content to better comprehend.

When monitoring progress toward goals, a learner is monitoring the fit between their learning results and previously set learning goals for the session. The aspects of the learning situation that are monitored during MPTG are the learner's domain knowledge, their expectations of results, and the learning goals. Closely related to time monitoring, MPTG is an essential monitoring activity that learners should use to stay "on-track" to the completion of the learning task. A learner may be able to generate several critical sub-goals for their learning task, but if they do not monitor the completion or incompleteness of these sub-goals, the sub-goal generation was inadequate. When a learner monitors the progress toward goals and realizes that he/she has only accomplished one out of four of their sub-goals in 90% of the time devoted to the learning task, a good self-regulator will revisit the remaining sub-goals and decide which is most important to pursue next.

In time monitoring, the learner is monitoring the task condition of time, with respect to their pre-existing learning goals. These learning goals can be either the global learning goal defined before engaging in the learning task, or sub-goals created by the learner during the learning episode. If the learner recognizes that very little time remains and few of the learning goals have been accomplished, they should make adaptations to the manner in which the material is being tackled. For example, if a learner has been reading a very long passage for several minutes and realizes that they have not accomplished the learning goals, a good self-regulator will begin scanning remaining material for information related to the goals not yet reached.

When learners engage in content evaluation, they are monitoring the appropriateness or inappropriateness of learning material they are currently reading, hearing, or viewing with regards to the overall learning goal or sub-goal they are currently pursuing. In contrast to content evaluation, evaluation of adequacy of content relates to the learner's assessment of the appropriateness of available learning content, rather than content currently being inspected. The aspects of the learning situations monitored in both of these processes are the learning resources and the learning goals. The learner should remain aware of whether learning goals and learning resources are complementary. If a learner evaluates a particular piece of material as particularly appropriate given their learning goal, they should direct more cognitive resources toward this material (or navigate toward this material), and persist in reading or inspecting the content in order to achieve this goal. Conversely, if a particular content is evaluated as inappropriate with respect to a learning goal, a good self-regulator will navigate away from (or not at all toward) this content to seek more appropriate material. In sum, these monitoring processes are based on studies examining the role of self-regulatory processes deployed by learners during learning with hypermedia.

The Deployment of Self-Regulatory Processes During Learning with Hypermedia

One of the purposes of treating self-regulated learning as an event, rather than an aptitude, is that we can capture, trace, and analyze the deployment of these processes during learning. It should be noted that the deployment of such processes unfolds both linearly and recursively during learning (based on Winne, 2001; Winne & Perry, 2000; Winne, Jamieson-Noel, & Muis, 2002). More specifically, the linear unfolding refers to the sequential deployment during learning. For example, a learner may set a goal and then enact a learning strategy such as taking notes and then assess the use of note-taking as not particularly helpful in facilitating their learning about the topic. This linear deployment of processes reflects a basic methodological assumption common to all on-line cognitive trace

methodologies including think-aloud protocols (i.e., learners' access to the contents of working memory and the sequential, linear enactment of both cognitive and behavioral self-regulatory processes; see Ericsson, 2006; Ericsson & Simon, 1993). In this section we highlight the theoretical, conceptual, and methodological importance of capturing and analyzing the deployment of self-regulatory processes during learning. We will accomplish this goal by providing evidence regarding the deployment of nearly three dozen self-regulatory processes during hypermedia learning, highlighting the top three self-regulatory processes used by learners across three developmental levels, and then focusing specifically on one participant.

A major issue relates to whether learners, in general, deploy the same amount and same types of self-regulatory processes throughout a learning task. Contemporary SRL models and frameworks (e.g., Pintrich, 2000; Schunk, 2005; Winne, 2001; Zimmerman, 2000, 2001) of SRL postulate that learners initially activate their prior knowledge, set goals, and plan their learning by coordinating their goals throughout the learning session. Despite the accuracy and plausibility of the enactment of such processes, the majority of research using these models has not treated SRL as an event and as such there is no published literature on the accuracy of such models based on SRL as an event, and for learning with CBLEs such as hypermedia. As such, our approach provides evidence regarding SRL as an event. We provide four examples in this section after we briefly describe our methodological approach.

Figure 17.1 illustrates the proportion of SRL moves of 132 middle-school, high-school, and college students' coded think-aloud protocols (based on a random sample of existing data from students learning with hypermedia). The *x*-axis is divided into four 10-minute episodes of the hypermedia learning task while the *y*-axis represents the proportion of SRL processes deployment based on aggregate SRL processes—planning, monitoring, learning strategies, handling task difficulties and demands, and interest. As can be seen, there is a difference in the proportion of SRL processes used during the hypermedia learning task.

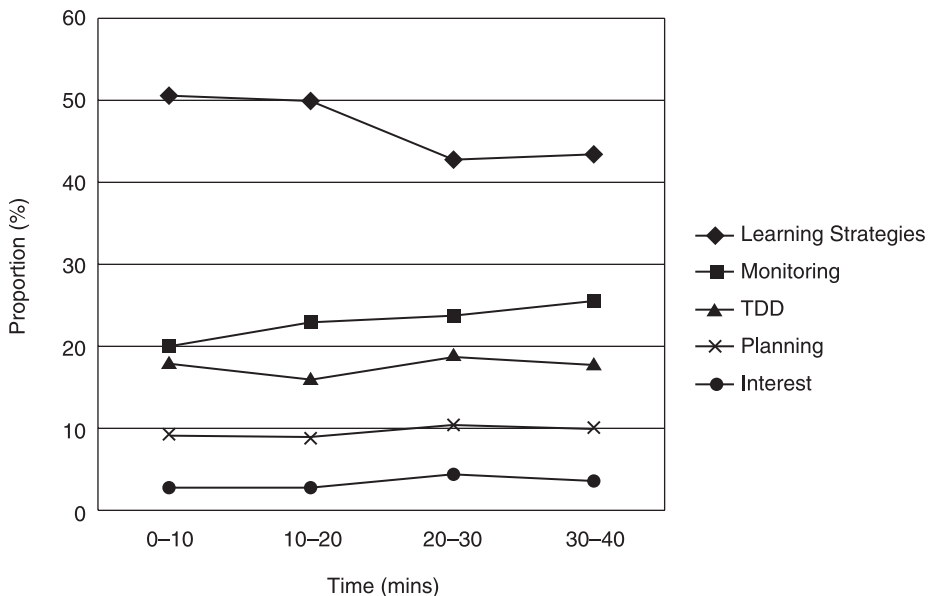


Figure 17.1 Proportion of SRL moves by time interval during a hypermedia learning task.

Note: TDD = Handling task difficulties and demands.

Overall, approximately 50% of the processes used by learners (at any point during the task) are learning strategies. This is then followed by monitoring processes (between 20% and 25%), handling task difficulties and demands (approx. 20%), planning (approx. 10%), and interest (approx. 5%). A few observations can be made (based upon this figure), including the fact that there is little fluctuation in the deployment of these processes during a 40-minute hypermedia learning task. The higher proportion of learning strategies and monitoring processes (in Figure 17.1) during the course of the hypermedia learning task is supported theoretically and conceptually by Winne and Hadwin’s (1998, 2008) model of SRL. This data can also be used to generate certain hypotheses regarding the deployment of SRL. For example, the slight “dip” in the proportion of learning strategies halfway through the task may be indicative of the low-prior-knowledge learners switching from a knowledge acquisition mode of learning (all they can about the circulatory system) to a knowledge-building mode of learning where monitoring processes (such as FOK) may now be more relevant (as seen in the slight rise in proportion of monitoring processes, see Figure 17.2). These hypotheses can be empirically tested.*

While this data is quite important in determining the consistency of SRL processes during learning with hypermedia, there is also a need to examine whether there are particular SRL processes that account for the shape of the lines seen in Figure 17.1. So, for each of the aggregated data points in Figure 17.1, we illustrate the three most frequently used SRL processes by the same sample. The results are presented in Figure 17.2, which presents the same information as in Figure 17.1, the only difference being that we inserted

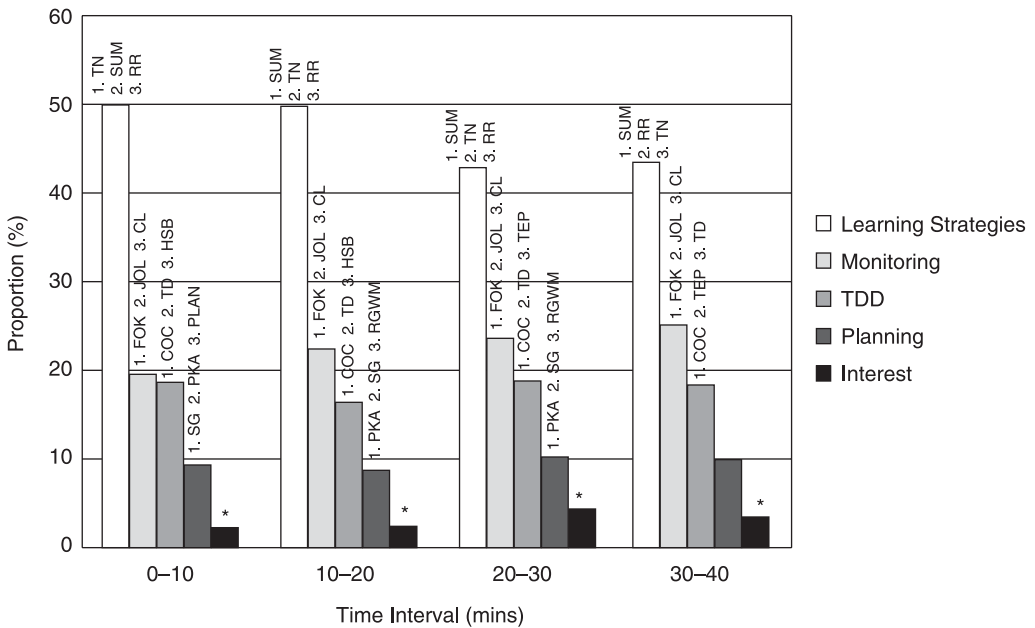


Figure 17.2 Three most frequently deployed SRL processes for each SRL cluster by time interval during a hypermedia learning task.

Note: TN = taking notes; SUM = summarizing; RR = re-reading; FOK = feeling of knowing; JOL = judgment of learning; CE = content evaluation; COC = control of context; TD = task difficulty; HSB = help-seeking behavior; SG= sub-goals; PKA = prior knowledge activation; PLAN = planning; TEP = time and effort planning; RGWM= recycling goals in working memory.

* The proportion of SRL moves at each time interval illustrated in Figure 17.1 is based on the relative sum of all participants’ self-regulatory processes across all SRL processes within each category, at each time interval.

the top three most frequently used SRL processes for each SRL cluster at each of the four 10-minute intervals. As can be seen, it is quite remarkable that the same three most frequently used SRL processes are taking notes, summarization, and re-reading, at each of these four time intervals. During the middle of the hypermedia learning task (i.e., between the 10–20 and 20–30 minute episodes) there is a switch between summarization, taking notes, and re-reading. This is then followed by another switch in the top three learning strategies—taking notes being the least used of the top three learning strategies. This finding provides some support for the hypothesis that learners will decrease their use of the learning strategy take notes toward the end of the learning session, instead turning to more summarization and re-reading of what has already been learned within the hypermedia environment.

Figure 17.2 illustrates that the three most frequently used monitoring processes are FOK, JOL, and CE. In addition, this sequence of use remains the same throughout the entire task. It is important to highlight that learners tend to monitor whether or not they are familiar with the content (by using FOK) by attempting to link it with prior knowledge. They also monitor their emerging understanding of the topic during the hypermedia learning task by engaging in JOL. Monitoring the adequacy of the multiple external representations (i.e., text, diagrams, animation) presented in the hypermedia environments is crucial in determining the appropriateness of each of these representations, vis-à-vis one's learning goals. These monitoring processes are used with remarkable consistency throughout the task, which highlights the importance of several monitoring processes related to a learner's prior knowledge, emerging understanding, and hypermedia content. We will not discuss planning, handling task difficulties and demands and interest due to space limitations.

A related question is whether we can more specifically forecast the percentage with which some SRL processes precede and follow each other. Table 17.2 shows the three most common preceding SRL processes for each of the monitoring processes. For example, out of all occurrences of FOK, 13% of the time the code was preceded by summarization, 12% by re-reading, and 12% by JOL. Occurrences of JOL were preceded by summarization 14% of the time, 10% of the time by prior knowledge activation, and 9% of the time by FOK. As for content evaluation, it was preceded by control of context 15% of the time, 12% by FOK, and 11% of the time by taking notes.

Table 17.3 shows the three most common SRL processes that follow each of the

Table 17.2 Most Common Preceding Moves Before Each Monitoring Process

<i>Monitoring process</i>	<i>Most common preceding move</i>	<i>Second most common preceding move</i>	<i>Third most common preceding move</i>
Feeling of Knowing (FOK)	SUM (13%)	RR (12%)	JOL (12%)
Judgment of Learning (JOL)	SUM (14%)	PKA (10%)	FOK (9%)
Monitoring Progress Towards Goals (MPTG)	SUM (12%)	TN (12%)	COC (11%)
Self-Questioning (SQ)	SQ (13%)	SUM (12%)	COC (10%)
Content Evaluation (CE)	COC (15%)	FOK (12%)	TN (11%)
Monitoring Use of Strategies (MUS)	TN (18%)	FOK (15%)	DRAW (12%)

* Normal text indicates learning strategies, shadow: planning, *italics*: monitoring, and underline: handling task difficulties and demands

Note: TN = taking notes; SUM = summarizing; SQ = self-question; COC = control of context; RR = re-reading; PKA = prior knowledge activation; FOK = feeling of knowing; JOL = judgment of learning; DRAW = drawing.

Table 17.3 Most Common Subsequent Moves Following Each Monitoring Process

Monitoring process	Most common subsequent move	Second most common subsequent move	Third most common subsequent move
Feeling of Knowing (FOK)	SUM (14%)	<u>COC (13%)</u>	FOK (9%)
Judgment of Learning (JOL)	FOK (17%)	SUM (9%)	RR (8%)
Monitoring Progress Towards Goals (MPTG)	<u>COC (14%)</u>	SG (10%)	PKA (10%)
Self-Questioning (SQ)	<u>SQ (13%)</u>	JOL (12%)	FOK (9%)
Content Evaluation (CE)	<u>COC (14%)</u>	SUM (12%)	FOK (9%)
Monitoring Use of Strategies (MUS)	TN (14%)	DRAW (11%)	<u>TEP (11%)</u>

* Normal text indicates learning strategies, shadow: planning, *italics*: monitoring, and underline: handling task difficulties and demand

Note: SUM = summarizing; RR = re-reading; FOK = feeling of knowing; COC = control of context; SQ = self-question; TN = taking notes; SG = sub-goals; JOL = judgment of learning; DRAW = drawing; PKA = prior knowledge activation; TEP = time and effort planning.

monitoring processes. For example, out of all occurrences of FOK, 14% of the time this process was followed by a summarization, 13% by control of context, and 9% by another FOK. In order for two codes of the same SRL process to occur within a transcript, another uncoded activity (e.g., reading content) must occur between the codes.

At a micro-level we can also analyze the deployment of self-regulatory processes. We also contributed to the much needed theoretically derived and empirically based research methods that characterize temporally unfolding patterns of engagement with tasks in terms of the phases and areas that constitute SRL (see Figure 17.3). Figure 17.3 illustrates the SRL trace data of a “high-jumper” (i.e., a learner who showed a significant mental model shift from pre-test to post-test) during a 40-minute hypermedia learning task (from Azevedo, Moos, Greene, Winters, & Cromley, 2008). The x-axis represents the number of SRL moves coded based on the think-alouds throughout the learning session while the y-axis represents each of the coded SRL processes. Blue vertical lines are used to delineate each 5-minute episode in the 40-minute learning task.

In addition, the same SRL process data can be used to calculate state change probability estimates for clusters of SRL processes. Table 17.4 illustrates the state change table depicting probabilities of learners’ subsequent self-regulatory moves based on a recent study with 82 college students (Azevedo et al., 2007). The table is read from the top to the side—for example, if the learner’s first move is planning, then there is a 0.43 chance

Table 17.4 State Change Table Depicting Probabilities of Learners’ Subsequent Self-Regulatory Moves

	<i>1st self-regulatory move</i>				
	<i>Planning</i>	<i>Monitoring</i>	<i>Strategy use</i>	<i>Task difficulty and demands</i>	<i>Interest</i>
<u><i>2nd self-regulatory move</i></u>					
Planning	0.09	0.10	0.08	0.10	0.13
Monitoring	0.29	0.32	0.17	0.21	0.26
Strategy use	0.43	0.38	0.62	0.48	0.41
Task difficulty and demands	0.17	0.16	0.11	0.19	0.09
Interest	0.02	0.03	0.02	0.02	0.11

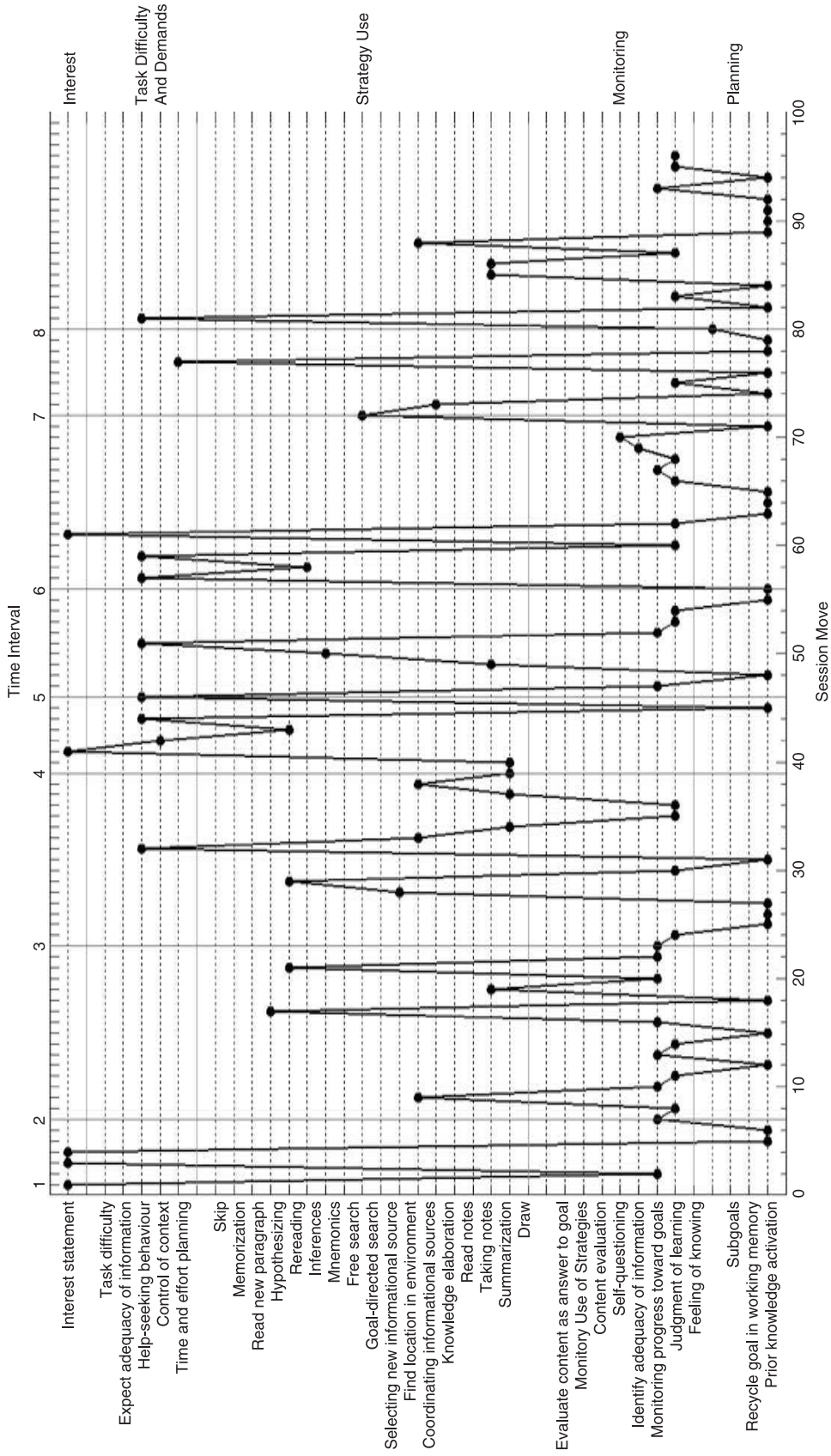


Figure 17.3 SRL trace data from a participant shifting from low mental model pre-test to high mental model post-test.

probability that their next move (reading off the right-hand side of the table) will be a strategy. Similarly, if the first move is a strategy then there is a 0.62 probability that the next move will also be a strategy. This table shows an interesting trend—regardless of what the first move is, the next most probable moves in descending order of magnitude are strategies, monitoring processes, planning activities, methods of handling task difficulties and demands, and then interest. These probability tables are useful in determining the granularity of assessing SRL processes and have tremendous utility in building adaptive hypermedia learning environments.

Research that compares learning patterns over time and reflects regulation per se in a well-defined instructional context (a single learner learning about a complex science topic with a hypermedia environment) is needed. There is a need to use multiple measurements and to triangulate these different data sources. This can lead to a better understanding of SRL and to theory-building in the area which can then serve as an empirical basis for the design of CBLEs. Recently, several researchers (Biswas et al., 2005; Witherspoon, Azevedo, Greene, Moos, & Baker, 2007) have designed CBLEs that are both authentic environments for studying, as well as for gathering trace data about self-regulated learning events during learning. We have adopted a similar approach, and are presently constructing a hypermedia environment designed to study the components of SRL.

Empirically-Based Guidelines for Monitoring During SRL with Hypermedia

The results of contemporary research on SRL and hypermedia have implications for the design of hypermedia environments intended to foster students' learning of complex and challenging science topics. At a global level, an adaptive hypermedia environment could be designed to deploy several key self-regulated learning processes such as planning (e.g., coordinating several sub-goals necessary to accomplish the overall learning task), monitoring aspects of one's cognitive architecture (e.g., through JOL and FOK) and the contents of the adaptive hypermedia system (e.g., identifying the adequacy of information sources) vis-à-vis one's progress (e.g., by monitoring progress toward goals), deploying effective learning strategies (e.g., summarizing, drawing, and taking notes), and handling task difficulties and demands (e.g., time and effort planning, controlling the context, and acknowledging and dealing with task difficulty) based upon an individual learner's behavior. Due to space limitation we focus specifically on monitoring processes.

The following prescriptive guidelines can be used to enhance learners' monitoring during learning with hypermedia:

- During learning of a science topic, learners could be prompted periodically to plan and activate their prior knowledge. This could be used as an anchor from which to build new understandings and elaborations based on the information presented in the hypermedia.
- Scaffolds could be designed to encourage a learner to engage in several metacognitive processes during learning. For example, the system could prompt a learner to engage FOK by asking them periodically (e.g., after reading a section of text or coordinating several information sources) to relate the information presented in the hypermedia environment to what they already know about the topic (i.e., knowledge elaboration).
- Static scaffolds could be embedded in the system to facilitate a learner's monitoring of their progress towards goals by using the planning net and sub-goals mentioned above and having the learner actively verify that he or she has learned about each of the sub-goals given the time remaining in the learning session.

- A learner's deployment of effective learning strategies could be scaffolded by providing on-line, embedded prompts. The system might also be able to detect the use of both effective and ineffective learning strategies, and provide prompts and feedback designed to discourage learners from using ineffective strategies.

Conclusion

In this chapter, we have provided a synthesis of existing research on learning with hypermedia by focusing on key cognitive and metacognitive processes related to self-regulation. There are several key theoretical, conceptual, methodological, and analytical issues that need to be addressed in order to advance our understanding of metacognitive monitoring and control processes related to hypermedia learning. Theoretically, there is a need to build a unified model of metacognition and self-regulated learning that incorporates key aspects of existing models, assumptions, processes, mechanisms, and phases. Such a process-oriented model would be ideal because it could be used to generate testable hypotheses regarding the facilitative and inhibitory nature of certain processes and underlying mechanisms and their impact on learning and performance. Conceptually, the field needs clarification on the multitude of divergent and overlapping constructs and processes across models to build a unified framework and model of metacognitive monitoring and control. Methodologically, the field needs to use multi-method studies and use current and emerging technological tools and sensors to capture and identify the dynamics of self-regulatory processes as they unfold linearly, recursively, and adaptively during learning. Furthermore, new statistical and analytical procedures need to be advanced in order to align and analyze temporal data from multiple sources such as eye-tracking, think-alouds, and facial expressions (denoting motivational and affective states) with existing measures (e.g., self-report measures) and methods. Lastly, the proposed issues would lead to theoretically-based and empirically-derived principles for designing advanced learning technologies aimed at detecting, monitoring, and fostering learners' cognitive and metacognitive self-regulatory processes.

Authors' Note

Please address all correspondence to: Roger Azevedo, University of Memphis, Department of Psychology and Institute for Intelligent Systems, 400 Innovation Drive, Memphis, TN, 38152. E-mail: razevedo@memphis.edu

The research presented in this chapter has been supported by funding from the National Science Foundation (Early Career Grant 0133346, 0633918, and 0731828) and a sub-contract from the Social Sciences and Humanities Research Council of Canada (SSHRC) awarded to the first author. We acknowledge the contributions of current and former members of the Cognition and Technology Laboratory (<http://azevedolab.autotutor.org/>) including Gwyneth Lewis, Shanna Smith, Moongee Jeon, Jeremiah Sullins, Emily Siler, Sarah Leonard, Andrew Trousdale, Jennifer Scott, Michael Cox, Dr Jennifer Cromley, Dr Daniel Moos, Dr Jeffrey Greene, and Fielding Winters. The authors thank Gwyneth Lewis and three reviewers for providing comments on an earlier version of this chapter.

References

- Ainley, M. & Patrick, L. (2006). Measuring self-regulated learning processes through tracking patterns of student interaction with achievement activities. *Educational Psychology Review*, 18(3), 267–286.

- Alexander, P. (1995). Superimposing a situation-specific and domain-specific perspective on an account of self-regulated learning. *Educational Psychologist*, 30(4), 189–193.
- Azevedo, R. (2002). Beyond intelligent tutoring systems: Computers as MetaCognitive tools to enhance learning? *Instructional Science*, 30(1), 31–45.
- Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist*, 40, 199–209.
- Azevedo, R. (2007). Understanding the complex nature of self-regulated learning processes in learning with computer-based learning environments: An introduction. *Metacognition and Learning*, 2(2/3), 57–65.
- Azevedo, R. (2008). The role of self-regulation in learning about science with hypermedia. In D. Robinson & G. Schraw (Eds.), *Recent innovations in educational technology that facilitate student learning* (pp. 127–156). Charlotte, NC: Information Age Publishing.
- Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, 29, 344–370.
- Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C., & Greene, J. A. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science*, 33, 381–412.
- Azevedo, R., Greene, J. A., & Moos, D. C. (2007). The effect of a human agent's external regulation upon college students' hypermedia learning. *Metacognition and Learning*, 2(2/3), 67–87.
- Azevedo, R., Guthrie, J. T., & Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research*, 30(1), 87–111.
- Azevedo, R. & Jacobson, M. (2008). Advances in scaffolding learning with hypertext and hypermedia: A summary and critical analysis. *Educational Technology Research & Development*, 56(1), 93–100.
- Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. C. (2008). Why is externally-regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development*, 56(1), 45–72.
- Benjamin, A. & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 73–94). New York: Taylor and Francis.
- Biswas, G., Leelawong, K., Schwartz, D., & TAGV (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19, 363–392.
- Boekaerts, M. & Cascallar, E. (2006). How far we have moved toward the integration of theory and practice in self-regulation. *Educational Psychology Review*, 18, 199–210.
- Boekaerts, M., Pintrich, P., & Zeidner, M. (2000). *Handbook of self-regulation*. San Diego, CA: Academic Press.
- Butler, D. & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245–281.
- Dillon, A. & Gabbard, R. (1998). Hypermedia as educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research*, 68(3), 322–349.
- Dillon, A. & Jobst, J. (2005). Multimedia learning with hypermedia. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 569–588). Cambridge, MA: Cambridge University Press.
- Dunlosky, J., Hertzog, C., Kennedy, M., & Thiede, K. (2005). The self-monitoring approach for effective learning. *Cognitive Technology*, 9, 4–11.
- Dunlosky, J. & Lipko, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, 16, 228–232.
- Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1, 3–14.
- Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness,

- R. R. Hoffman, & P. J. Feltovich (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 223–242). Cambridge, MA: Cambridge University Press.
- Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Cambridge, MA: MIT Press.
- Graesser, A., McNamara, D., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AuthTutor and iSTRAT. *Educational Psychologist*, 40(4), 225–234.
- Greene, J. A. & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research*, 77(3), 334–372.
- Greene, J. A. & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of sophisticated mental models. *Contemporary Educational Psychology*, 34, 18–29.
- Hadwin, A., Nesbit, J., Jamieson-Noel, D, Code, J., & Winne, P. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, 2, 107–124.
- Hadwin, A., Winne, P., & Stockley, D. (2001). Context moderates students' self-reports about how they study. *Journal of Educational Psychology*, 93, 477–487.
- Hegarty, M., Narayanan, H., & Freitas, P. (2002). Understanding machines from multimedia and hypermedia presentations. In J. T. Otero & J. León (Eds.), *Psychology of science text comprehension* (pp. 357–384). Mahwah, NJ: Erlbaum.
- Jacobson, M. (2008). A design framework for educational hypermedia systems: Theory, research, and learning emerging scientific conceptual perspectives. *Educational Technology Research & Development*, 56, 5–28.
- Jacobson, M. J. & Azevedo, R. (2008). Scaffolding learning with hypertext and hypermedia: Theoretical, empirical, and design issues. *Educational Technology Research and Development*, 56(1), 1–3.
- Koriat, A. & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103, 490–517.
- Lawless, K. & Kulikowich, J. (1996). Understanding hypertext navigation through cluster analysis. *Journal of Educational Computing Research*, 14, 385–399.
- Liu, M. (2004). Examining the performance and attitudes of sixth graders during their use of a problem-based hypermedia learning environment. *Computers in Human Behavior*, 20, 357–379.
- McNamara, D. & Shapiro, A. (2005). Strategies for helping learners create coherence from text and hypertext. *Journal of Educational Computing Research*, 33, 1–29.
- Metcalfe, J. & Dunlosky, J. (2008). Metamemory. In H. Roediger (Ed.), *Cognitive psychology of memory* (Vol. 2). Oxford: Elsevier.
- Moos, D. C. & Azevedo, R. (2008a). Self-regulated learning with hypermedia: The role of prior knowledge. *Contemporary Educational Psychology*, 33, 270–298.
- Moos, D. C. & Azevedo, R. (2008b). Exploring the fluctuation of motivation and use of self-regulatory processes during learning with hypermedia. *Instructional Science*, 36, 203–231.
- Moos, D. C. & Azevedo, R. (in press). Learning with computer-based learning environments: A literature review of computer self-efficacy. *Review of Educational Research*.
- Neiderhauser, D. (2008). Educational hypertext. In M. Spector, D. Merrill, J. van Merriënboer, & M. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 199–209). New York: Taylor and Francis.
- Nelson, T. & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). San Diego, CA: Academic Press.
- Pintrich, P. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.
- Schunk, D. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist*, 40(2), 85–94.
- Schwartz, N., Andersen, C., Hong, N., Howard, B., & McGee, S. (2004). The influence of

- metacognitive skills on learners' memory of information in a hypermedia environment. *Journal of Educational Computing Research*, 31(1), 77–93.
- Shapiro, A. (1999). The relationship between prior knowledge and interactive overviews during hypermedia-aided learning. *Journal of Educational Computing Research*, 20(2), 143–167.
- Shapiro, A. & Niederhauser, D. (2004). Learning from hypertext: Research issues and findings. In D. Jonassen (Ed.), *Handbook of research for education communications and technology* (2nd ed.). Mahwah, NJ: Erlbaum.
- Tergan, S. (1997a). Conceptual and methodological shortcomings in hypertext hypermedia design and research. *Journal of Educational Computing Research*, 16, 209–235.
- Tergan, S. (1997b). Misleading theoretical assumptions in hypertext/hypermedia research. *Journal of Educational Multimedia and Hypermedia*, 6(3–4), 257–283.
- Thiede, K. & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 1024–1037.
- Van Overschelde, J. (2008). Metacognition: Knowing about knowing. Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 47–71). New York: Taylor and Francis.
- Veenman, M. (2007). The assessment and instruction of self-regulation in computer-based environments: A discussion. *Metacognition and Learning*, 2, 177–183.
- Vollmeyer, R. & Rheinberg, F. (2006). Motivational effects on self-regulated learning with different tasks. *Educational Psychology Review*, 18, 239–253.
- Winne, P. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Mahwah, NJ: Erlbaum.
- Winne, P. (2005). Key issues on modeling and applying research on self-regulated learning. *Applied Psychology: An International Review*, 54(2), 232–238.
- Winne, P. & Hadwin, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ: Erlbaum.
- Winne, P. & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). New York: Taylor & Francis.
- Winne, P., Jamieson-Noel, D. L., & Muis, K. (2002). Methodological issues and advances in researching tactics, strategies, and self-regulated learning. In P. R. Pintrich & M. L. Maehr (Eds.), *Advances in motivation and achievement: New directions in measures and methods* (pp. 121–155). Greenwich, CT: JAI Press.
- Winne, P. & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego, CA: Academic Press.
- Witherspoon, A. M., Azevedo, R., & D'Mello, S. K. (2008). Transitions in self-regulatory behavior in self-regulated and externally-regulated learning episodes. Proceedings of the 9th international conference on Intelligent Tutoring Systems.
- Witherspoon, A., Azevedo, R., Greene, J. A., Moos, D. C., & Baker, S. (2007). The dynamic nature of self-regulatory behavior in self-regulated learning and externally-regulated learning episodes. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (pp. 179–186). Amsterdam, The Netherlands: IOS Press.
- Zeidner, M., Boekaerts, M., & Pintrich, P. (2000). Self-regulation: Directions and challenges for future research. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 750–768). San Diego, CA: Academic Press.
- Zimmerman, B. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.
- Zimmerman, B. (2001). Theories of self-regulated learning and academic achievement: An overview

and analysis. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 1–37). Mahwah, NJ: Erlbaum.

Zimmerman, B. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *Review of Educational Research*, 45, 166–183.

Zimmerman, B. & Schunk, D. (2001). *Self-regulated learning and academic achievement* (2nd ed.). Mahwah, NJ: Erlbaum.

18 Interactive Metacognition

Monitoring and Regulating a Teachable Agent

*Daniel L. Schwartz, Catherine Chase,
Doris B. Chin, Marily Oppizzo, and
Henry Kwong*
School of Education, Stanford University

Sandra Okita
Teachers College, Columbia University

*Gautam Biswas, Rod Roscoe, Hogyong
Jeong, and John Wagster*
Department of Engineering, Vanderbilt University

Metacognition involves monitoring and regulating thought processes to make sure they are working as effectively as possible (Brown, 1987; Flavell, 1976; Winne, 2001). Good teachers are highly metacognitive (Lin, Schwartz, & Hatano, 2005; Sherin, 2002). They reflect on their expertise and instruction, and they refine their pedagogy accordingly. Good teachers are also metacognitive in a less conventional sense of the term. They monitor student understanding and they regulate the processes that the students use to learn and solve problems (Shulman, 1987). Thus, good teachers apply metacognition to other people's thoughts. The proposal of this chapter is that asking children to teach and apply metacognition to others can help them learn both content knowledge and metacognitive skills. A strong version of this proposal, consistent with Vygostky (1978), would be that metacognition develops first on the external plane by monitoring others, and then turns inward to self-monitoring. The chapter does not test this claim. Instead, it shows that having students teach a specially designed computer agent leads to metacognitive behaviors that increase content learning and hint at improving metacognition more generally.

To differentiate self-directed metacognition and other-directed metacognition, we term the latter "interactive metacognition." Learning-by-teaching is an instructional method that is high on interactive metacognition—good tutors anticipate, monitor, regulate, and more generally interact with their tutees' cognition. Research on learning-by-teaching has found that teaching another person is an effective way to learn. For instance, when people prepare to teach pupils to take a test, they learn more compared to when they prepare to take the test themselves (Annis, 1983; Bargh & Schul, 1980; Biswas, Schwartz, Bransford, et al., 2001; cf. Renkl, 1995). Moreover, during the act of teaching, tutors learn by clarifying the confusions of their tutees (Craig, Sullins, Witherspoon, & Gholson, 2006; Palincsar & Brown, 1984; Uretsi, 2000) and by engaging in reflective knowledge building (Roscoe & Chi, 2008). Interestingly, when tutors slip into "lecturing mode" and no longer engage in interactive metacognition, they learn less (Chi, Roy, & Hausmann, 2008; Fuchs, Fuchs, Bentz, Phillips, & Hamlett, 1994; Graesser, Person, & Magliano, 1995; Roscoe & Chi, 2007).

The interactive quality of other-directed metacognition can help resolve two psychological challenges: balancing the dual-task demands of metacognition and rallying the motivation to engage in metacognition. Metacognition puts a dual-task load on working memory. During metacognition, people need (1) to think their problem-solving thoughts, and they simultaneously need (2) to monitor and regulate their thinking about those thoughts. When learning or problem solving becomes difficult, there can be less free capacity for metacognition. For example, when first learning to drive a car with a manual transmission, people may be less likely to monitor their knowledge of the cars behind them. Teaching can help alleviate the dual-task demand of metacognition. The tutee has the responsibility of problem solving, which frees up resources for the tutor's metacognition. Gelman and Meck (1983), for example, found that young children could monitor errors in adult counting better than their own counting, when the counting task reached the edge of the children's abilities (cf. Markman, 1977). In this case, interactive metacognition was a form of distributed cognition (King, 1998; Kirsch, 1996), where the adult took on the burden of problem solving and the child took on the burden of monitoring that problem solving.

The distribution of tasks in interactive metacognition can help students improve their own metacognition, because they can focus on monitoring and regulating cognition per se. For example, in a series of studies by Okita (2008), elementary-school children learned tricks for mentally solving complex arithmetic problems. In half of the cases, students practiced problem solving on their own. In the other half of the cases, students took turns. On one turn they would try to solve a problem, and on the next turn they would monitor a computer agent solving a problem. The children had to stop the agent if they thought there was a mistake. Students who monitored the agent demonstrated a U-shaped curve in their own problem solving. When first monitoring the agent, students subsequently became slower and less accurate in their own problem solving. Over time, however, the children sped up and became more accurate compared to students who never monitored the agent. Presumably, by monitoring the agent, the students were learning to monitor themselves, which caused a temporary drop in efficiency, but a better payoff in the long run, because they improved their own cognition.

The second challenge of metacognition is motivational. Because metacognition takes extra work, people will tend to "get by" if they can, rather than take the extra cognitive effort needed to go beyond "good enough" (Martin & Schwartz, in press). Students often skim readings, because they think it is not worth checking their understanding. Teachers, however, are responsible for their students' performance, not to mention their own display of competence. This increase in responsibility can motivate teachers to engage in metacognition, which may be one reason that tutors learn more when preparing to teach than simply studying for themselves (e.g., Annis, 1983).

This chapter reviews research on Teachable Agents to demonstrate that it is possible to use computer learning environments to produce the cognitive and motivational benefits of interactive metacognition. Teachable Agents are an instance of a pedagogical computer agent (e.g., Baylor, 2000; Dillenbourg, Mendelsohn, & Schneider, 1994; Johnson, Rickel, & Lester, 2000). Pedagogical agents are computer characters that have been designed to help students learn. With Teachable Agents, students learn by explicitly teaching an intelligent computer agent. The chapter begins with an introduction of the Teachable Agent, "Betty's Brain," followed by a description of how Betty elicits interactive metacognitive behaviors. The chapter then shows that teaching Betty improves children's content learning and their abilities to use the same sort of reasoning as Betty. Finally, the chapter examines students' learning choices to determine whether they begin to internalize interactive metacognition.

A Technology for Applying Interactive Metacognition

This section explains how the Teachable Agent (TA) software naturally engages metacognition during learning. Betty’s Brain, the TA shown in Figure 18.1 and the focus of this chapter, was designed for knowledge domains where qualitative causal chains are a useful structural abstraction (e.g., the life sciences). Students teach Betty by creating a concept map of nodes connected by qualitative causal links; for example, “burning fossil fuels” increases “carbon dioxide.” Betty can answer questions based on how she was taught. For instance, Betty includes a simple query feature. Using basic artificial intelligence reasoning techniques, Betty animates her reasoning process as she answers questions (for technical details, see Biswas, Leelawong, Schwartz, Vye, & TAG-V, 2005). In Figure 18.1, Betty uses the map she was taught to answer the query, “What happens to ‘heat radiation’ if ‘garbage’ increases?” Students can trace their agents’ reasoning, and then remediate their agents’ knowledge (and their own) if necessary. As described below, there are many feedback features that help students monitor their agent’s understanding. A version of the Betty’s Brain environment and classroom management tools can be found at <aaalab.stanford.edu/bettylogin.html>. Betty is not meant to be the only means of instruction, but rather, she provides a way to help students organize and reason about the content they have learned through other classroom lessons.

In reality, when students work with Betty, they are programming in a high-level, graphical language. However, Betty’s ability to draw inferences gives the appearance of sentient behavior. Betty also comes with narratives and graphical elements to help support the

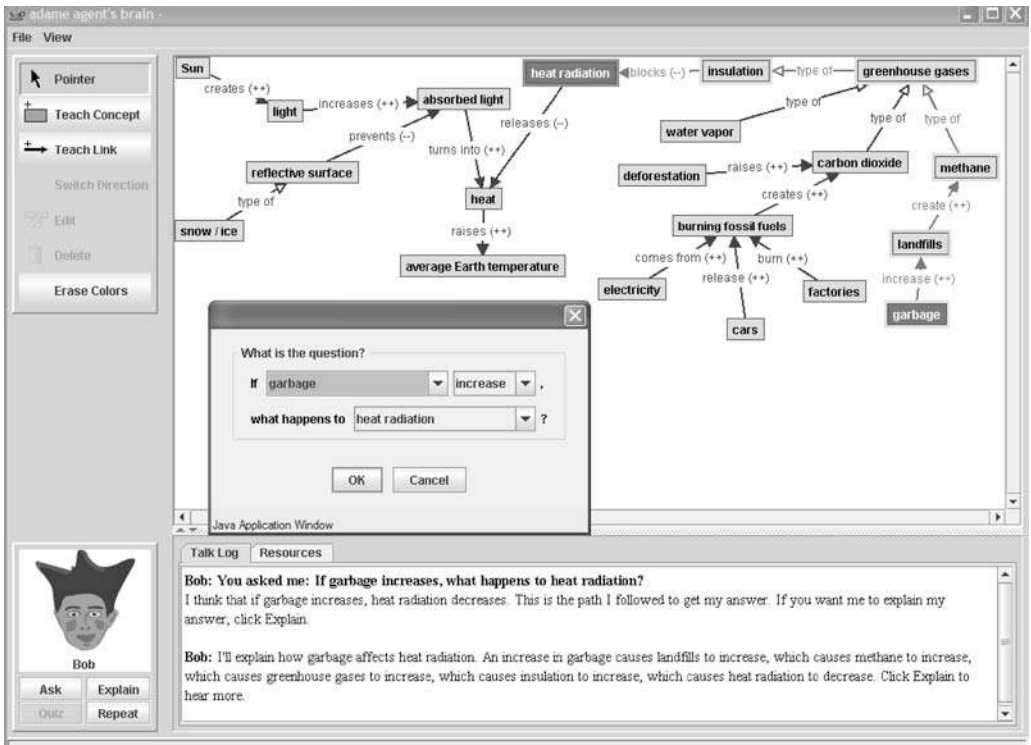


Figure 18.1 The Teachable Agent Named Betty. The student has (a) named his agent “Bob” instead of Betty, (b) customized Bob’s look, (c) taught Bob about global warming, and (d) asked Bob what happens to heat radiation if garbage increases.

mindset of teaching; for example, students can customize their agent's appearance and give it a name. ("Betty's Brain" is the name of the software, not a student's specific agent.) Betty can also take quizzes, play games, and even comment on her own knowledge. Ideally, a TA enlists students' social imagination so they will engage in the processes of monitoring and regulating their agents' knowledge.

A key element of Betty is that she externalizes thought processes. Betty literally makes her thinking visible. Thus, students are applying metacognition to the agents' thinking and because that thinking is visible, it is an easily accessible format.

Monitoring One's Own Thoughts in an Agent

For students to practice metacognition on their agent, they need to view Betty as exhibiting cognitive processes. This section shows that students do treat their agent as sentient, which leads them to take responsibility for monitoring and regulating their agents' knowledge. It then shows that Betty's knowledge is a fair representation of the students' own knowledge, which shortens the distance between monitoring the agent and monitoring themselves.

The Agent Elicits Interactive Metacognitive Behaviors

When programming and debugging their agents, students are also monitoring and regulating their agents' knowledge and reasoning. A study with fifth-graders demonstrated that students treat their agents as having and using knowledge (Chase, Chin, Oppezzo, & Schwartz, in press). By this age, children know the computer is not really alive, but they suspend disbelief enough to treat the computer as possessing knowledge and feelings (e.g., Reeves and Nass, 1996; Turkle, 2005). Students monitor their agents' failures and share responsibility, which leads them to revise their own understanding so they can teach better.

The study used the Triple-A-Challenge Gameshow, which is an environment where multiple TAs, each taught by a different student, can interact and compete with one another (Figure 18.2). Students can log on from home to teach their agents, chat with other students, and eventually have their agents play in a game. During game play, (1) the game host poses questions to the agents; (2) the students wager on whether their agent will answer correctly; (3) the agents answer based on what the students taught them; (4) the host reveals the correct answer; and finally, (5) wager points are awarded. In addition to boosting engagement, the wagering feature was intended to lead students to think through how their agent would answer the question, thereby monitoring their agent's understanding. The Triple-A-Challenge Gameshow was developed to make homework more interactive, social, and fun. In this study, however, the focus was on student attitudes towards Betty during game play, and students were videotaped as they worked alone.

The study included two conditions. In both, students received a text passage on the mechanisms that sustain a fever, and they taught their TA about these concepts. The treatment difference occurred when playing the Gameshow. In the TA condition, the agents answered six questions, and the graphical character in the Gameshow represented the student's agent. In the Student condition, the students answered the questions, and the character represented the student. To capture students' thoughts and feelings towards the agent, students in both groups thought aloud.

In the TA condition, students treated their agents as having cognitive states. Students' attributions of cognitive states were coded as being directed to themselves, their agents, or both. Examples of self-attributions include, "It's kind of confusing to me," "I have a really good memory," and "No, actually, I don't know." Examples of agent-attributions include,

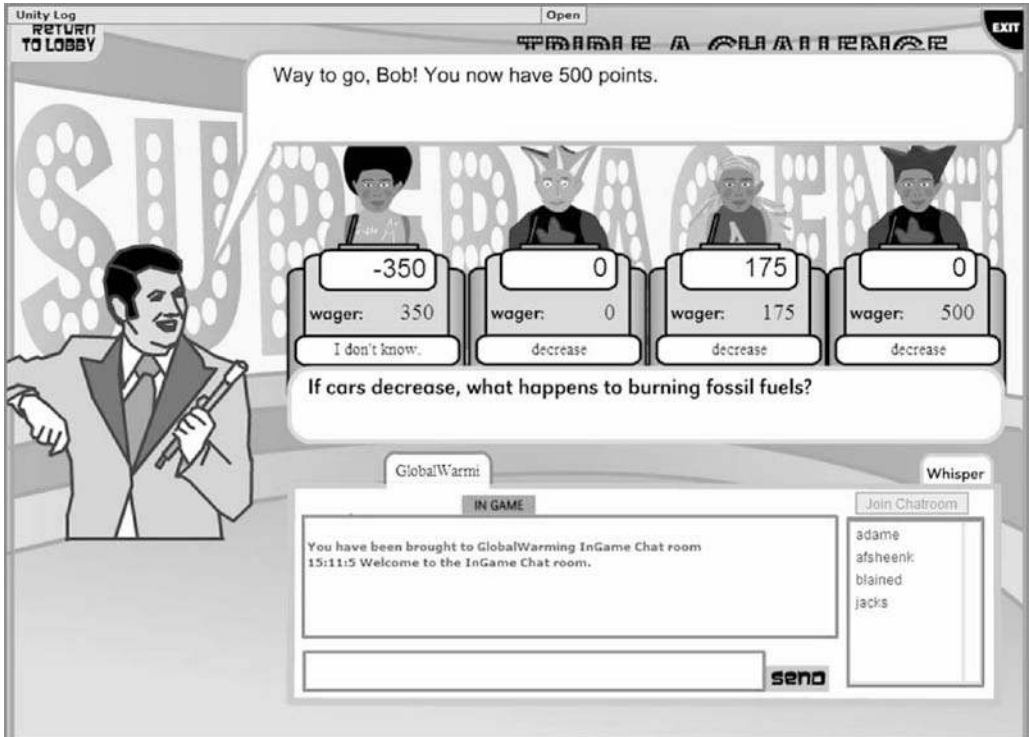


Figure 18.2 The Triple-A-Challenge Gameshow. Students log on for homework. After teaching their agents, the agents play with one another. A host asks questions of each agent. Students wager on whether they think their agent will give the right answer. The agents respond based on what the students taught them. There is a chat window so students can communicate with one another during the game.

“He doesn’t know it,” and, “He knows if shivering increases. . . .” Sometimes, a single statement could include both self- and agent-attributions; for example, “I’m pretty sure he knows this one,” and, “I guess I’m smarter than him.”

During game play, students in both treatments made about two cognitive state attributions per question. For the TA condition, over two-thirds of these attributions were towards the agent or a combination of agent and student. Thus, students treated the agent as a cognitive entity, and in fact, they sometimes confused who was doing the thinking, as in the case of one boy who stated, “ ’cause I don’t . . . ’cause he doesn’t know it.”

The TA students also took an “intentional stance” (Dennett, 1989) towards their agents, by apportioning responsibility to the agent for success and failure. They could have behaved as though all successes and failures were theirs, because the agent is simply a program that generates answers from a map the student had created, but they did not. Table 18.1 indicates the number of attribution-of-credit statements made in response to successful and unsuccessful answers. Examples of success attributions include, “I’m glad I got it right” (self), “He got it right!” (agent), or, “We got it!” (both). Examples of failure attributions include, “I didn’t teach him right” (self), “He said *large* increase when it was only increase” (agent), or, “Guess we were wrong” (both).

As the table shows, students in the TA condition liberally attributed responsibility to the agent. Importantly, the TA condition exhibited more attention to failure, which is a key component of monitoring (e.g., Zimmerman & Kitsantas, 2002). They made nearly

Table 18.1 Average Number of Attributions to Success and Failure (and Standard Errors of the Mean)

Condition	Attributions when succeeded				Attributions when failed			
	Self	Agent	Both	Total	Self	Agent	Both	Total
TA answers	.17(0.12)	.27(0.12)	0.0(0.0)	.44(0.16)	.54(0.13)	.47(0.21)	.66(0.19)	1.67(0.28)*
Student answers	.53(0.10)	n/a	n/a	.53(0.10)	.65(0.22)	n/a	n/a	.65(0.22)

Note: * $p < 0.05$ – comparison of condition means

three times as many remarks in a failure situation relative to the Student condition. The attributions were spread across themselves and their agents. In addition to acknowledging failure, they often made remarks about flaws in their teaching such as, “Whoa. I really need to teach him more.” Thus, at least by the verbal record, the TA condition led the students to monitor and acknowledge errors more frequently than the Student condition.

The study also demonstrated that the students were sufficiently motivated by teaching to engage in the extra work that metacognition often entails. After completing the round of game play, students were told the next round would be more difficult. They were given the opportunity to revise their maps and re-read the passage in preparation. While all the children in the TA condition chose to go back and prepare for the next round, only two-thirds of the Student condition prepared. Of those who did prepare, the TA students spent significantly more time at it. The protocol data from the game play help indicate one possible reason. The Student condition exhibited nearly zero negative responses to failure (e.g., “Ouch!”). Given an unsuccessful answer, the Student condition averaged 0.02 negative affective responses. In contrast, the TA condition averaged 0.62 expressions of negative affect. Much of this negative affect was regarding their agent’s feelings. For example, one student said, “Poor Diokiki . . . I’m sorry Diokiki,” when his agent, Diokiki, answered a question incorrectly. The TA students felt responsibility for their agents’ failures, which may have caused them to spend more time preparing for the next round of game play (which they did not actually play due to time constraints).

Overall, these data indicate that the children treated their agents as if they were sentient, which had subsequent effects on their learning behaviors. In reality, the children were “playing pretend.” They knew their agent was not a sentient being. Regardless, their play involved the important features of metacognition—thinking about mental states and processes, noticing and taking responsibility for mistakes, and experiencing sufficient affect that it is worth the effort to do something about the mistakes when given a chance to revise. Working with another, in this case an agent one has taught, can lead to more metacognitive behaviors than completing a task oneself.

The Agent’s Knowledge Reflects the Student’s Knowledge

Schoenfeld (1987), discussing the importance of monitoring, states that “. . . the key to effective self-regulation is being able to accurately self-assess what is known and not known.” With Betty, students are assessing what their agent does and does not know. The agent’s knowledge is a reflection of their own knowledge, so that working with the agent indirectly entails working on an externalized version of their own knowledge. This was demonstrated by correlating the test scores of the students and their agents.

Betty can be automatically tested on the complete population of questions in a concept map. By using a hidden expert map that generates the correct answers, the program can successively test Betty on all possible questions of the form, “If node <X> increases, what

happens to node <Y>?” The results produce an APQ Index (all possible questions) that summarizes the overall test performance of the TA.

A study with 30 sixth-grade students compared the agents’ APQ indices with how well students did on their own paper-and-pencil tests. Students completed three cumulative units by teaching their agents about global warming and climate change. At the end of each unit, the agents were tested to derive an APQ Index, and students took a short-answer paper-and-pencil test. In the paper-and-pencil test, half of the items comprised TA-like questions, in the sense that they depended on causal chaining and nodes in Betty’s map. The other half comprised Non-TA questions, in the sense that they depended on content that was not captured in Betty’s nodes. The Non-TA questions helped to determine whether Betty correlated with student knowledge more broadly, and not just questions that Betty could answer.

Table 18.2 indicates that the TA scores were positively correlated with students’ test scores. Overall, these correlations are better than the correlations between students’ own scores on the TA-like questions and the Non-TA questions for each unit test (see Note in Table 18.2). Thus, the APQ Index correlated better with student performance on the TA-like and Non-TA questions than these two types of paper-and-pencil items correlated with each other. (The low correlations for Test 3 are due to a badly worded TA-like question.) Conceivably, with further development and evaluation, it will be possible to test agents instead of students, thereby saving valuable instructional time.

The correlation of student and agent performance indicates that when students monitor their agent’s knowledge, for example by asking it a question, they are likely to be monitoring a fair externalization of their own knowledge. This helps to dissolve the gap between self and other, so that the children’s task of working with the agent is a proxy for the task of reflecting upon their own knowledge.

Adopting the Cognition of the Agent

Given that students treat the TA as exhibiting mental states and the TA reflects the student’s knowledge, the next question is whether these have any effect on student learning. Ideally, by monitoring another’s cognition, one can pick up the other person’s style of reasoning. Siegler (1995), for example, found that young children learned number conservation more effectively when prompted to explain the experimenter’s reasoning rather than their own. Betty reasons by making inferences along causal chains. When students teach Betty, they learn to simulate her causal reasoning for themselves.

Learning to simulate Betty’s cognition about a situation is different from learning to

Table 18.2 Correlations between Students’ Agents and Students’ Test Scores

Agent	Student test scores								
	All questions			TA-like questions			Non-TA questions		
	Test 1	Test 2	Test 3	Test 1	Test 2	Test 3	Test 1	Test 2	Test 3
APQ Index									
Test 1	60**	–	–	0.51**	–	–	0.56**	–	–
Test 2	–	0.66**	–	–	0.47*	–	–	0.66**	–
Test 3	–	–	0.34	–	–	0.12	–	–	0.48*

Note: ** $p < 0.01$; * $p < 0.05$. Correlations of TA-like and non-TA questions: $r = 0.47$ (test 1); $r = 0.46$ (test 2); and $r = 0.14$ (test 3).

simulate the situation itself. When people reason about a situation itself, they often create a mental model that helps them imagine the behavior of the situation and make predictions (Gentner & Stevens, 1983; Glenberg, Gutierrez, Levin, Japuntich, & Kascak, 2004; Zwaan & Radvansky, 1998). For example, when reasoning about how gears work, people can create and simulate an internal image of the gears to solve problems (Schwartz & Black, 1996). To run their mental model, people imagine the forces and movements of the gears, and they observe the resulting behaviors in their mind’s eye. With Betty, students create a mental model of the agent’s reasoning. So, rather than simulating forces and spatial displacements, the students learn to simulate chains of declarative reasoning. This way, Betty’s cognition becomes internalized as a way of reasoning for the student.

Relevant data come from a study where two classes of sixth-graders learned about global warming. Over three weeks, students learned the mechanisms of the greenhouse effect, the causes of greenhouse gases, and finally, the effects of global warming. Both classes completed hands-on activities, saw film clips, received lectures, and completed relevant readings. At regular points, students were asked to create concept maps to organize their learning, and they all learned how to model causal relations using a concept map. The difference was that one class was assigned to the Betty condition; these students used the Betty software to make their concept maps. Figure 18.3 shows a finished “expert” version of a map created on the Betty system. The other class was assigned to the Self condition; these students used Inspiration®, a popular, commercial concept-mapping program.

Students in both conditions received multiple opportunities for feedback with an important difference. In the Betty condition, agents answered the questions, and the feedback was directed towards the agents. In the Self condition, the students answered the questions, and the feedback was directed towards them. This difference occurred across several feedback technologies. For example, the agents took quizzes or the students took quizzes. For homework, the agents answered questions in the Gameshow or the students answered the questions in the Gameshow. Thus, the main difference between conditions was that in the Betty condition, the learning interactions revolved around the task of teaching and monitoring the agent, whereas in the Self condition, the learning interactions

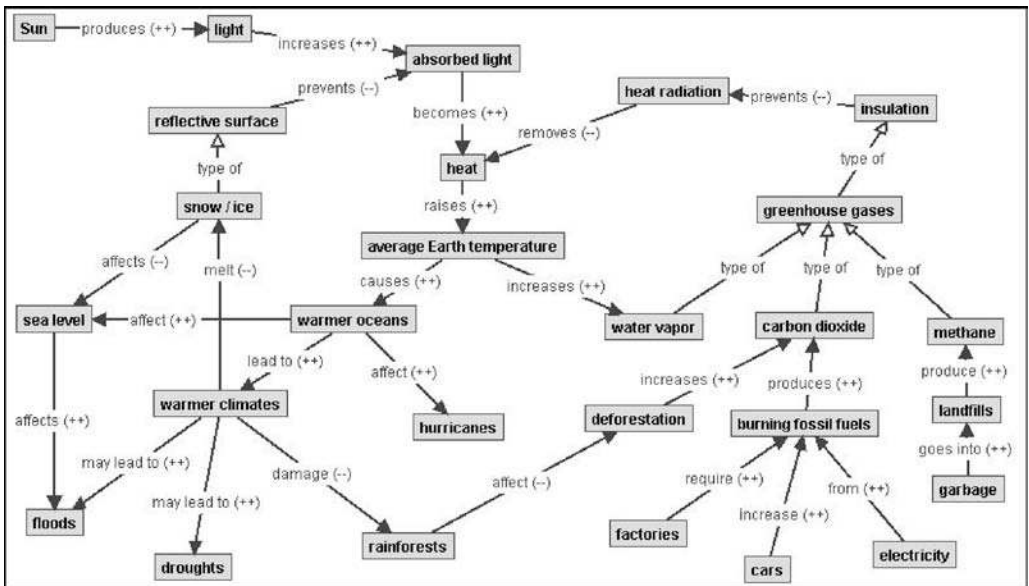


Figure 18.3 Target knowledge organization for global warming curriculum.

revolved around the task of creating a concept map and answering questions and monitoring oneself.

The students in the Betty condition adopted Betty’s reasoning style. After each unit—mechanisms, causes, effects—all the students completed short-answer paper-and-pencil tests. The tests included questions that required short, medium, or long chains of causal inference. An example of a short-chain question involved answering why warmer oceans increase sea level. An example of a long-chain question involved detailing the causal chain that spans from an increase in factories to the effects on polar bears. Figure 18.4 shows that over time the Betty students separated from the Self students in their abilities to complete longer chains of inference. After the first unit, the two groups overlapped, with the Betty students showing a very modest advantage for the longer inferences. After the second unit, the TA students showed a strong advantage for the medium-length inferences. By the final unit, the TA students showed an advantage for short, medium, and long inferences.

This study used intact classes, so the results are promissory rather than conclusive. Nevertheless, the steady improvement in length of causal inference is exactly what one would expect the Betty software to yield, because this is what the agent’s reasoning models and enforces. The interactive metacognition of teaching and monitoring Betty’s reasoning and her accuracy helped students internalize her style of thinking, which in this case is a positive outcome because her reasoning involved causal chaining.

Regulating Cognition for the Agent

In addition to monitoring cognition, metacognition involves taking steps to guide cognition, or as it is often termed, “regulating” cognition (Azevedo & Hadwin, 2005; Brown, 1987; Butler & Winne, 1995; Pintrich, 2002; Schraw, Crippen, & Hartley, 2006). Regulating another can help students learn to regulate for themselves.

Thus far, Betty’s features supported monitoring, but there were few features to help students decide what to do if they detected a problem. For example, one student’s agent

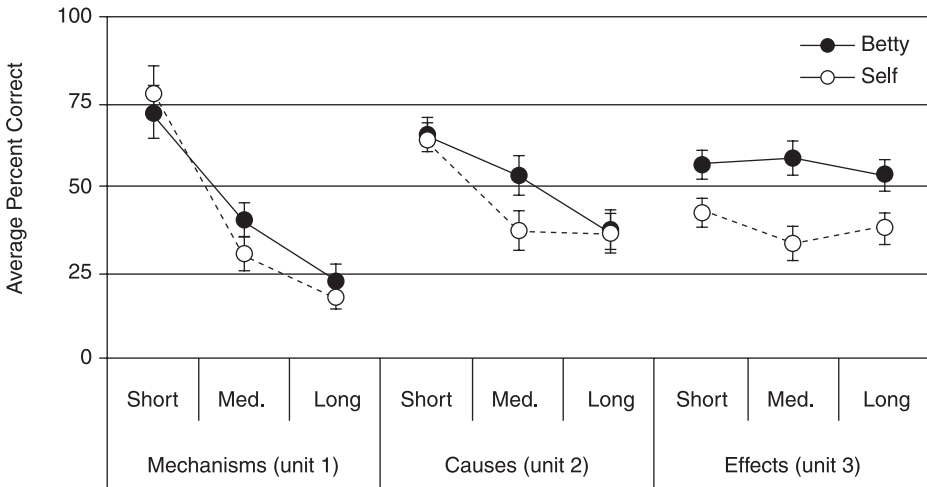


Figure 18.4 Effects of Betty versus Self. Each test included questions that depended on short, medium, or long chains of causal inference to answer correctly. With more experience across the lesson units, Betty students showed an increasing advantage for longer causal inferences. The Self condition used the concept mapping software Inspiration® instead of Betty.

was performing poorly in the Gameshow, and the student did not know how to fix it. The Gameshow was not designed to address this situation directly. Fortunately, another student used the Gameshow's chat feature to provide support, "Dude, the answer is right there in the reading assignment!"

To help students learn to self-regulate their thinking, Betty comes in a self-regulated learning (SRL) version. For example, when students add incorrect concepts or links, Betty can spontaneously reason and remark that the answer she is deriving does not seem to make sense. This prompts students to reflect on what they have just taught Betty and to appreciate the value of checking understanding. SRL Betty also includes Mr Davis, a mentor agent shown in Figure 18.5. Mr Davis complements the teaching narrative, because he grades Betty's quizzes and gives her feedback on her performance. This feedback is in the form of motivational support (e.g., "Betty, your quiz scores are improving"), as well as strategies to help the students improve Betty's knowledge (e.g., "Betty, ask your teacher to look up the resources on quiz questions that you have got wrong . . .").

SRL Betty implements regulation goals specified by Zimmerman's (1989) list of self-regulation strategies. The SRL system monitors for specific patterns of interaction, and when found, Betty or Mr Davis provides relevant suggestions (also see Jeong et al., 2008). Table 18.3 provides a sample of triggering patterns and responses used by the SRL system; there are many more than those shown in Table 18.3.

In sum, SRL Betty is an adaptive tutoring system, except that students are the tutors, and

The screenshot shows the 'Betty's Brain' interface. On the left is a toolbar with options like 'Pointer', 'Teach Concept', 'Teach Link', 'Edit', and 'Delete'. Below the toolbar is a portrait of Mr. Davis with the text 'Ask Mr. Davis'. The main area displays a concept map with nodes like 'Nitrosomonas bacteria', 'nitrites', 'Nitrobacter bacteria', 'anemonia', 'fish', 'animal waste', 'snails', and 'macroinvertebrates', connected by arrows with labels like 'produce', 'consumed by', 'need', 'broken down into', 'generate', 'eat', 'help', 'feed', 'hel', 'proc', and 'breakth'. A window titled 'Mr. Davis is grading Betty's quiz. Please wait...' is overlaid on the map, showing a door labeled 'Mr. Davis' and a sign that says 'Testing. Do not disturb.' At the bottom, a 'Quiz 3' results table is visible.

Question	Quiz Answer	Current Answer
1. If fish increase, what happens to algae?	An increase	--
2. If algae decrease, what happens to fish?	A decrease	--
3. If fish increase, what happens to nitrites?	A large increase	--
4. If anemonia decrease, what happens to Nitrobacter bacteria?	A decrease	--

Figure 18.5 Adding SRL to Betty's Brain. The student has submitted Betty to take a quiz given by Mr Davis, and the results are shown in the bottom panel. Mr Davis and Betty provide tips and encouragement for engaging in metacognitive behaviors.

Table 18.3 Sample of Interactive Patterns and Responses by the SRL System

<i>Regulation goal</i>	<i>Pattern description</i>	<i>Betty response</i>	<i>Mr Davis response</i>
MONITORING THROUGH EXPLANATION	Multiple requests for Betty to give an answer but no request for explanation	Let's see, you have asked me a lot of questions, but you have not asked for my explanations lately. Please make me explain my answers so you will know if I really understand.	Without asking Betty to explain her answers, you may not know whether she really understands the chain of events that you have been trying to teach her. Click on the Explain button to see if she explains her answer correctly.
SELF-ASSESSMENT	Repeated quiz request but no updates have been made to the map	Are you sure I understand what you taught me? Please ask me some questions to make sure I got it right. I won't take the quiz otherwise. Thanks for teaching me about rivers!	You have not taught Betty anything new. Please, spend some time teaching her new links and concepts and make sure she understands by asking her questions. Then she can take the quiz again. If you need help learning new things, check the resources.
TRACKING PROGRESS	The most recent quiz score is significantly worse than the previous	I would really like to do better. Please check the resources, teach me, and make sure I understand by asking me questions that are on the quiz. My explanation will help you find out why I am making mistakes in my answers. Also, be sure to check out the new tips from Mr. Davis.	Betty did well on the last quiz. What happened this time? Maybe you should try rereading some of the resources and asking Betty more questions so that you can make sure she understands the material.
SETTING LEARNING GOALS	Betty is asked a question that she cannot answer for the second time	I just don't know the relationships yet, maybe we should ask Mr. Davis what we need to learn.	I've seen this kind of difficulty with teaching some of my own students in the past. You should try looking for missing link connections or links that are in the wrong direction.

the system adapts to target metacognitive needs specifically. The metacognitive support is integrated into the teaching narrative through computer characters that take the initiative to express opinions, make requests, and provide relevant prompts to encourage further interactive metacognition. In the following, the first sub-section shows that SRL support helps students learn science content. The second sub-section introduces a new machine learning methodology for analyzing student choices. The methodology is used to identify high-level interaction patterns that indicate metacognitive strategies. The methodology is then used to evaluate whether students developed metacognitive strategies that they continued to use on their own after the SRL features were turned off.

Self-Regulation Support Improves Student Learning

The self-regulation support in SRL Betty helps students learn science content better. Fifty-one fifth-grade students learned about interdependence in a river ecosystem with a special focus on the relations between fish, macroinvertebrates, plants, and algae. The students

worked over seven class periods starting with the food chain, then photosynthesis and respiration, and finally the waste cycle. To help the students learn, there were quizzes and reading resources built into the system. (In the Gameshow studies described earlier in the chapter, the students received the nodes, and their task was to determine the links. In this study, the students had to use the readings to decide which nodes to create in their maps.)

The study had three conditions: Regulated Learning by Teaching (RT); Learning by Teaching (LT); and Intelligent Coaching (IC). The RT condition used SRL Betty. As usual, students could query Betty to answer a question of the form “if X increases/decreases what happens to Y?” Betty would trace the path of her reasoning and give her final answer. If students further asked Betty to explain her answer, she would trace through a subset of the links, and color code the corresponding nodes to indicate increases and decreases. For the new SRL components, Betty and Mr Davis could also provide metacognitive tips and behaviors, some of which are included in Table 18.3. For example, when students submitted Betty to take a quiz, Mr Davis provided metacognitive hints about resources and steps the students could use to teach Betty better. Mr Davis did not dictate specific changes to Betty’s knowledge, for example, to add a particular concept or change a link. Instead, he suggested strategies for improving Betty’s knowledge (e.g., “Check if Betty understands after you have taught her something new”).

In the LT condition, students worked with Betty and the mentor agent, but without the SRL support. Betty did not provide prompts for regulating how she was taught, and Mr Davis provided direct instructions for how to fix the concept map after a quiz. For example, Mr Davis might tell students “to consider how macroinvertebrates might affect algae, add an appropriate link.”

The Intelligent Coach (IC) condition was identical to the LT condition, except there was no teaching cover story. Students used the software to make concept maps of their own knowledge. Instead of asking Betty to answer a question, students could ask Mr Davis to answer a question using the concept map or to explain how the map gave a certain answer. Thus, students received the same information and animations as in the LT condition, except they thought it was their map that Mr Davis was analyzing instead of Betty’s thinking.

In addition to the initial seven days of learning, the study included a second learning phase that measured transfer. Six weeks after completing the river ecosystem unit, students left their original conditions to spend five class periods learning about the land-based nitrogen cycle. All the students worked with a basic Betty version. There were on-line reading resources; Betty could answer questions; and students could check how well Betty did on quizzes. However, there was no extra support, such as how to improve Betty’s map or their teaching. The logic of this phase was that if students had developed good metacognitive strategies, they would be more prepared to learn the new content on their own (Bransford & Schwartz, 1999).

The students’ final concepts maps from the main and transfer phases were scored for the inclusion of correct nodes and links based on the reading materials. Table 18.4 holds the average scores. Overall, both conditions that involved teaching did better than the Intelligent Coach condition, and there were no interactions with the phase of the study. This means that the Learning-by-Teaching condition did better than the Intelligent Coach condition, even though the only treatment difference between these two conditions was whether students thought they were teaching and monitoring Betty (LT) or being monitored by Mr Davis (IC). This result reaffirms the findings from the preceding global warming study using a tighter experimental design. If students believe they are teaching an agent, it leads to superior learning even when they are using the same concept-mapping tool and receiving equivalent feedback.

Table 18.4 Average Concept Map Scores at the end of the Main Treatment (River Ecosystems) and the Transfer Treatment (Land Nitrogen Cycle)

Condition	Study phase			
	Main learning (1 st)		Transfer for learning (2 nd)	
	Map Score		Map Score	
	M	(SE)	M	(SE)
(RT) ¹ Regulated Teaching	31.8 ^{3, 6}	(1.5)	32.6 ⁴	(2.9)
(LT) ² Learning-by-Teaching	25.8	(1.6)	31.8 ⁵	(3.0)
(IC) Intelligent Coach	22.4	(1.5)	22.6	(2.9)

Note: Overall treatment means greater than IC: ¹ $p < 0.01$; ² $p < 0.05$. Post hoc comparisons for each study phase – Greater than IC: ³ $p < 0.001$; ⁴ $p < 0.05$; ⁵ $p < 0.1$. Greater than LT: ⁶ $p < 0.05$

In a separate study not reported here, an Intelligent Coaching condition included self-regulated learning support, similar to the Regulated Teaching condition. (Mr Davis gave prompts for how to improve the concept map by consulting resources, checking the map by asking queries, etc.). In that study, the IC+Regulated support condition did no better than an IC condition, whereas the RT condition did. So, despite similar levels of metacognitive prompting, the prompting was more effective when directed towards monitoring and regulating one’s agent. This result also supports the basic proposition that teaching effectively engages metacognitive behaviors, even compared to being told to use those metacognitive behaviors for oneself.

Post hoc analyses of the main learning phase indicates that the extra metacognitive support of the RT treatment led to better initial learning than the LT condition in which students did not receive any guidance on regulation. However, once students lost the extra support in the transfer phase, they performed about the same as the LT students. By these data, self-regulation support helped students learn when it was available, but it is not clear that the extra support yielded lasting metacognitive skills compared to only teaching Betty. As described next, however, there were some modest differences in how the RT students went about learning in the transfer phase, even though these did not translate into significant learning differences.

Adopting Metacognitive Learning Choices from an Agent

Metacognition, besides helping people think more clearly, can also help people make choices about how to use learning resources in their environment. For example, to study for the California Bar exam, many students order the BAR/BRI materials (www.barbri.com). These materials comprise nearly a cubic meter of readings, reviews, outlines, practice tests, videotapes, as well as live local lectures, workshops and on-line tutorials. Across the materials, the content is highly redundant. Rather than plowing through all the materials, these well-educated adults often choose the presentation format and activities that they feel suit their learning needs and preferences for a particular topic. Their learning is driven by their choices of what, when, and how to learn. Outside of classrooms that exert strict control, this is often the case. People make choices that determine their learning. For younger students, metacognitive instruction should help children learn to make effective learning choices.

This section introduces a new data mining methodology for examining learning choices.

The goal is to be able to identify choice patterns that reflect effective metacognition. Ideally, once these patterns have been identified, adaptive technologies can monitor for these patterns and take appropriate actions. This is a useful endeavor, because current adaptive computer systems depend on strict corridors of instruction in which students can make few choices (except in the unrelated sense of choosing an answer to a problem). If students do not have chances to make choices during learning, it is hard to see how they can develop the metacognition to make effective learning choices. If the following methodologies and others in the future are successful, it will be possible to use more choice-filled learning environments, like virtual worlds, without sacrificing the benefits of adaptive technologies for helping students to improve.

To explore the data mining methodology, it was applied to the log files from the preceding study. The question was whether the methodology could help reveal whether the RT students exhibited unique patterns of learning choices during the initial learning phase when the metacognitive support was in play, and whether these patterns carried over to the transfer phase when the support was removed. That is, did the students in the RT condition internalize the metacognitive support so they exhibited effective metacognitive patterns once the support was removed?

In the study, students could make a number of choices about which activities to pursue. Table 18.5 summarizes the possibilities. For example, one student read the resources, and then made a number of edits to the map. Afterwards, the student submitted the map to a quiz, made some more edits, and then asked a pair of questions of the map. In raw form, the sequence can be overwhelming: R → M → M → M → A → M → M → M → M → M → M → M → M → M → Q → M → M → A → A.

To make sense of these complex choice sequences, a new data mining methodology analyzed the log files (Li & Biswas, 2002; Jeong & Biswas, 2008). The methodology automated the derivation of a Hidden Markov Model (HMM). An HMM model represents the probabilities of transitioning between different “aggregated” choice states (Rabiner, 1989). An aggregated choice state represents common choice patterns that comprise sequences of individual choices to transition from one activity to another. HMM is useful for identifying high-level choice patterns, much in the way that factor analysis is useful for identifying clusters of survey items that reflect a common underlying psychological property.

The HMM analysis generated three choice patterns that could be interpreted as increasing in metacognitive sophistication: Basic Map Building; Map Probing; and Map Tracing. The Basic Map Building pattern involves editing the map, submitting the map for a quiz, and occasionally referring to the reading resources. It reflects a basic and important metacognitive strategy. While students work on their maps, they check the map with a quiz to see if there are errors and occasionally look back at the readings.

Table 18.5 Possible Choices of Activities in SRL Betty System

<i>Activity name</i>	<i>Student actions</i>
Edit Map (M)	Adding, modifying, or deleting concepts and links.
Resource Access (R)	Accessing the reading resources.
Request Quiz (Q)	Submitting map to take a quiz.
Ask Query (A)	Asking Betty/Mentor to use map to answer a question.
Request Explanation (E)	Asking Betty/Mentor to explain an answer to a query.
Continue Explanation (C)	Asking for a more detailed explanation.

In the Map Probing pattern, students edit their maps, and then they ask a question of their map to check for specific relations between two concepts (e.g., if fish increase, what happens to algae?). This pattern exhibits a more proactive, conceptually driven metacognitive strategy. Students are checking specific causal chains rather than relying on the quiz to identify errors, and students need to formulate their own questions to check their maps.

Finally, the Map Tracing pattern captures when students ask Betty or Mr Davis (depending on the system) to explain the steps that led to an answer. When Betty or Mr Davis initially answers a question during Map Probing, the agents only state the answer and show the paths they followed. To see whether a specific link within the path produced an increase or decrease, students have to request an explanation. (Map Tracing can only occur after Map Probing.) Decomposing Betty’s explanations is particularly useful when maps become complex, and there are multiple paths between two concepts. Map Tracing is a sophisticated metacognitive strategy, because it involves decomposing a chain of reasoning step-by-step, even after the answer has been generated in Map Probing.

Figure 18.6 shows the complete set of transitional probabilities from one state to another

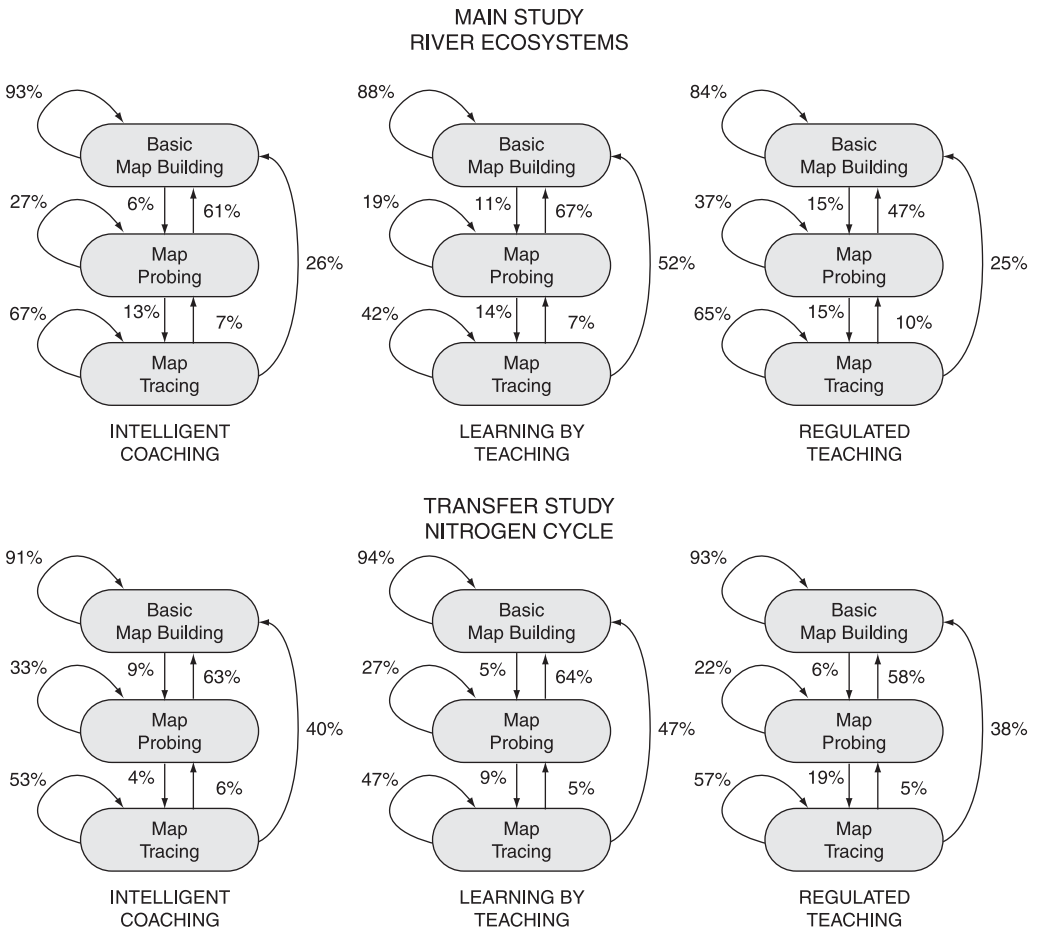


Figure 18.6 Transitional probabilities between aggregated choice states. Each state, derived through Hidden Markov Model statistical learning, represents a pattern of choices that create a common cluster of learning activities. The numbers beside the arrows indicate the probability that students would transition from one state to another.

another broken out by condition and phase of the study. The figure is complex, so the following discussion will narrow the focus to Map Tracing.

Multiplying the transition probabilities yields a rough estimate of the proportion of time students spent in a specific activity state. This is important, because just looking at a single transition can be misleading. For example, in the main phase of the study, the Intelligent Coach (IC) and Regulated Teaching (RT) conditions transitioned from Map Probing into Map Tracing at nearly the same rate. Nevertheless, the IC condition spent much less time Map Tracing. The IC students rarely transitioned from Map Building into Map Probing, and Map Probing is a necessary precursor to Map Tracing.

In the first phase of the study, students in all three conditions spent a significant proportion of their time in Basic Map Building. However, the RT students more frequently transitioned into the Map Probing and Map Tracing patterns. Their version of the software included two features to make this happen. First, Betty would not take a quiz if students had not checked her reasoning by asking her a question. This forced students to enter the Map Probing activity. Second, Betty and Mr. Davis suggested that the students ask Betty to explain her reasoning, so the students could trace her reasoning and look for errors. As a result, the proportions of effort spent in Map Probing and Tracing were twice as great for the RT condition compared to the other two conditions. Presumably, this contributed to the RT condition's superior content learning in the main phase of the study (see Table 18.4).

The metacognitive strategies practiced in the initial learning phase transferred somewhat when students had to learn the nitrogen cycle on their own. At transfer, when all students had to learn the nitrogen cycle without any special feedback or tips, the differences between conditions were much smaller. However, there was a telling difference that involved transitions into Map Tracing. The RT students, who had received regulation support, were twice as likely as the LT students to use Map Tracing. And the LT students who had taught Betty were twice as likely to use Map Tracing as the IC students. As ratios, the differences are quite large, though in terms of absolute amount of time spent Map Tracing, they are relatively small. Nevertheless, the strategic use of Map Tracing can help students get to the heart of key chains of reasoning. These differences may help explain why the LT and RT treatments learned more at post-test. These students were more likely to check how their agent was reaching its conclusion, which conceivably could have caused the superior learning.

At this point, it is only tentative that the self-regulation support in Betty affected students' learning at transfer via the learning choices they made. This HMM analysis aggregated across students and sessions within a condition. Thus, it is not possible to do statistical tests. Deriving patterns through HMM is a new approach to understanding students' metacognitive learning choices, and it is still being developed. The main promise of analyzing these patterns is that it can help improve the design of interactive, choice-filled environments for learning. By identifying better and worse interactive patterns for learning, it should be possible to design computer systems to identify those patterns in real-time and provide adaptive prompts to (a) move students away from ineffective metacognitive patterns, and (b) encourage them to use effective patterns. Thus, an important next step will be to correlate choice patterns with specific learning outcomes, so it is possible to determine which choice patterns do indeed lead to better learning.

Conclusion

The chapter's leading proposal is that teaching another person, or in this case an agent, can engage productive metacognitive behaviors. This interactive metacognition can lead to

better learning, and ideally, if given sufficient practice, students will eventually turn the metacognition inward.

The first empirical section demonstrated that students do take their agent's behavior as cognitive in nature, and that the agent's reasoning is correlated with the students' own knowledge. Thus, when students work with their agent, they are engaging in metacognition. It is interactive metacognition directed towards another. The second empirical section demonstrated that monitoring an agent can lead to better learning, because students internalize the agent's style of reasoning. In the final empirical section, the Teachable Agent was enhanced to include support to help children regulate their learning choices. Again, the results indicated that working with a Teachable Agent led to superior content learning, especially with the extra metacognitive support in place. Moreover, students who taught an agent made a near transfer to learn a new topic several weeks later.

An analysis of students' learning choices indicated that the students who had taught agents exhibited a more varied repertoire of choices for improving their learning. They also exhibited some modest evidence of transferring these metacognitive skills by choosing to check intermediate steps within a longer chain of inference.

It is informative to contrast Betty with other technologies designed as objects-to-think-with (Turkle, 2007). Papert (1980), for example, proposed that the programming language Logo would improve children's abilities to plan. Logo involved programming the movement of a graphical "turtle" on the computer screen. Evidence did not support the claim that Logo supported planning (Pea & Kurland, 1984). One reason might be that students had to plan the behavior of the turtle, but the logical flow of the program did not resemble human planning itself. For example, the standard programming construct of a "do-loop" involves iterating through a cycle and incrementing a variable until a criterion is reached. The execution logic of this plan does not resemble many human versions of establishing and managing a plan. Therefore, programming in Logo is an interactive task, but it is not a task where one interacts with external mental states or processes. In contrast, the way Betty reasons through causal chains is similar enough to human reasoning that programming Betty can be treated as working with her mental states. Students can internalize her cognitive structure, and eventually turn their thinking about her cognitive structures into thinking about their own.

Acknowledgments

This material is based upon work supported by the National Science Foundation under grants EHR-0634044, EHR-0633856, SLC-0354453, and by the Department of Education under grant IES R305H060089. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the granting agencies.

References

- Annis, L. (1983). The processes and effects of peer tutoring. *Human Learning*, 2, 39–47.
- Azevedo, R. & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition—Implications for the design of computer-based scaffolds. *Instructional Science*, 33, 367–379.
- Bargh, J. A. & Y. Schul (1980). On the Cognitive Benefits of Teaching. *Journal of Educational Psychology* 72, 593–604.
- Baylor, A. (2000). Beyond butlers: Intelligent agents as mentors. *Journal of Educational Computing Research*, 22, 373–382.

- Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & TAG-V. (2005). Learning By Teaching: A New Agent Paradigm for Educational Software. *Applied Artificial Intelligence*, 19, 363–392.
- Biswas, G., Schwartz, D. L., Bransford, J. D., & The Teachable Agents Group at Vanderbilt (2001). Technology Support for Complex Problem Solving. In K. D. Forbus & P. J. Feltoch (Eds.), *Smart Machines in Education: The coming revolution in educational technology* (pp. 71–97). Menlo Park, CA: AAAI Press.
- Bransford, J. D. & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education*, 24, 61–101. Washington DC: American Educational Research Association.
- Brown, A. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms, In F. E. Weinert & R. H. Kluwq (Eds.), *Metacognition, motivation and understanding*. New Jersey: Lawrence Erlbaum Associates
- Butler, D. & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245–281.
- Chase, C., Chin, D. B., Oppezzo, M., & Schwartz, D. L. (in press). Teachable agents and the protégé effect: Increasing the effort towards learning. *Education and Technology*.
- Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32, 301–341.
- Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level-reasoning-question effect: The role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction*, 24, 565–591.
- Dennett, D. (1989). *The intentional stance*. Cambridge, MA: MIT Press.
- Dillenbourg, P., Mendelsohn, P., & Schneider, D. (1994). The distribution of pedagogical roles in a multiagent learning environment. In R. Lewis & P. Mendelsohn (Eds.), *Lessons from learning* (pp. 199–216). Elsevier.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence*. New Jersey: Lawrence Erlbaum.
- Fuchs, L., Fuchs, D., Bentz, J., Phillips, N., & Hamlett, C. (1994). The nature of student interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, 31, 75–103.
- Gelman, R. & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition*, 13, 343–359.
- Gentner, D. & Stevens, A. (Eds.) (1983). *Mental Models*. Hillsdale, NJ: Erlbaum.
- Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, 96, 424–436.
- Graesser, A. C., Person, N., & Magliano, J. (1995). Collaborative dialog patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychologist*, 9, 359–387.
- Jeong, H. & Biswas, G. (2008). Mining Student Behavior Models in Learning-by-Teaching Environments. *Proceedings of the First International Conference on Educational Data Mining* (pp. 127–136), Montreal, Canada.
- Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. L. (2008). Using Hidden Markov Models to Characterize Student Behavior Patterns in Computer-based Learning-by-Teaching Environments. In B. Woolf, et al. (Eds.), *Intelligent Tutoring Systems: 9th International Conference, Montreal Canada, LNCS (vol. 5091)*, pp. 614–625.
- Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11, 47–78.
- King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. *Educational Psychology Review*, 10, 57–74.
- Kirsch, D. (1996). Adapting the environment instead of oneself. *Adaptive Behavior*, 4, 415–452.
- Li, C. & Biswas, G. (2002). A Bayesian Approach for Learning Hidden Markov Models from Data. *Special issue on Markov Chain and Hidden Markov Models, Scientific Programming*, 10, 201–219.

- Lin, X. D., Schwartz, D. L., & Hatano, G. (2005). Towards teacher's adaptive metacognition. *Educational Psychologist, 40*, 245–256.
- Markman, E. (1977). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development, 48*, 986–992.
- Martin, L. & Schwartz, D. L. (in press). Prospective adaptation in the use of representational tools. *Cognition and Instruction*.
- Okita, S. Y. (2008). Learn Wisdom by the Folly of Others: Children Learning to Self Correct by Monitoring the Reasoning of Projective Pedagogical Agents (Doctoral dissertation, Stanford University, 2008). Dissertation Abstracts International ProQuest.
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 2*, 117–175.
- Papert, S. A. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.
- Pea, R. D. & Kurland, D. M. (1984). On the cognitive effects of learning computer programming. *New Ideas Psychology, 2*, 137–168.
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice, 41*, 219–225.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*, 257–286.
- Reeves, B. & Nass, C. (1996). *The media equation*. New York: Cambridge University Press.
- Renkl, A. (1995). Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results. *Learning and Instruction, 5*, 21–36.
- Roscoe, R. & Chi, M. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*, 534–574.
- Roscoe, R. D. & Chi, M. (2008). Tutor learning: The role of instructional explaining and responding to questions. *Instructional Science, 36*, 321–350.
- Schoenfeld, A. H. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp.189–215). Hillsdale, NJ: Erlbaum.
- Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111–139.
- Schwartz, D. L. & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science, 20*, 457–497.
- Sherin, M. G. (2002). When teaching becomes learning. *Cognition and Instruction, 20*, 119–150.
- Siegler, R. S. (1995). How does change occur? A microgenetic study of number conservation. *Cognitive Psychology, 28*, 225–273.
- Turkle, S. (2005). *The second self: Computers and the human spirit, twentieth anniversary edition*. Cambridge, MA: MIT Press.
- Turkle, S. (Ed.) (2007). *Evocative objects: Things we think with*. Cambridge, MA: MIT Press.
- Uretsi, J. A. R. (2000). Should I teach my computer peer? Some issues in teaching a learning companion. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Intelligent Tutoring Systems* (pp. 103–112). Berlin: Springer-Verlag.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Soubberman, Eds. and Trans.). Cambridge, MA: Harvard University Press.
- Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Mahwah, NJ: Erlbaum.
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated learning. *Journal of Educational Psychology, 81*, 329–339.
- Zimmerman, B. J. & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*, 660–668.
- Zwaan R. A. & Radvansky G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185.

Part IX

Tutoring

19 Meta-Knowledge in Tutoring

Arthur C. Graesser and Sidney D'Mello

University of Memphis

Natalie Person

Rhodes College

It is important to have a chapter in this volume on one-to-one tutoring because it is one of the most effective methods of helping students learn. Tutoring provides the opportunity for an intense “meeting of the minds” between a student and a person with subject matter expertise. Tutoring is the standard panacea when students are failing to meet expected grades and anxiety reverberates among the students, parents, teachers, principals, and school systems. Anxiety turns to panic when a school is not meeting the standards of a high stakes test, when a prize athlete may be cut from a team, and when a student runs the risk of losing a scholarship. A wealthy family might end up paying \$200 per hour for an accomplished tutor to save a son or daughter.

It is very easy to justify the use of tutors from the standpoint of learning gains. Meta-analyses show learning gains from typical, non-expert human tutors of approximately 0.4 sigma (effect size in standard deviation units) compared to classroom controls and other suitable controls (Cohen, Kulik, & Kulik, 1982). Non-expert tutors are para-professionals, cross-aged tutors (i.e., students who are older than the tutee), or same-age peers who have had little or no tutor training and have modest subject-matter expertise. Collaborative peer tutoring shows an effect size advantage of 0.2 to 0.9 sigma (Johnson & Johnson, 1992; Mathes & Fuchs, 1994; Slavin, 1990; Topping, 1996), and appears to be slightly lower than older non-expert human tutors (Pilkington & Parker-Jones, 1996). The tutors in these same-age and cross-age collaborations tend to learn more than the tutees (Cohen et al., 1982; Mathes & Fuchs, 1994; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003), presumably because of the added study, effort, and initiative in taking on the role as tutor. Peer tutoring is a low-cost effective solution because expert tutors are expensive and hard to find. In contrast, there have not been many systematic studies on learning gains from expert tutors because they are expensive, they are difficult to recruit in research projects, and tutors tend to stay in the tutoring profession for a short amount of time. However, available studies show effect sizes of 0.8 to 2.0 (Bloom, 1984; Chi, Roy, & Hausmann, 2008; Person, Lehman, & Ozburn, 2007; VanLehn et al., 2007), which is presumably higher than other forms of tutoring.

The obvious question to ask is why is tutoring effective in promoting learning? This question has been investigated by researchers in education, cognitive science, and discourse processing in recent years. One approach to answering this question is to conduct a program of research or to report meta-analyses that relate learning gains with characteristics of the subject matter, tutee, tutor, and general structure of the tutoring session. There is evidence, for example, that (a) learning gains tend to be higher for well-structured, precise domains (mathematics, physics) than for ill-structured domains, (b) that learning gains from tutors are more pronounced for tutees who start out with comparatively lower amounts of knowledge and skill, (c) that the quality of tutor training is much more important than the quantity of training, and (d) that a tutoring session shows more

benefits when there are particular pedagogical activities (Cohen et al., 1982; Fuchs, Fuchs, Bentz, Phillips, & Hamlett, 1994; King, Staffieri, & Adelgais, 1998; Mathes & Fuchs, 1994; Rohrbeck et al., 2003). A second approach is to perform a very detailed analysis of the tutoring session structure, tasks, curriculum content, discourse, actions, and cognitive activities manifested in the sessions and to speculate how these might account for the advantages of tutoring (Chi et al., 2008; Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Graesser & Person, 1994; Graesser, Person, & Magliano, 1995; Hacker & Graesser, 2007; Lepper, Drake, & O'Donnell-Johnson, 1997; McArthur, Stasz, & Zmuidzinas, 1990; Merrill, Reiser, Merrill, & Landes, 1995; Person & Graesser, 1999; 2003; Person, Kreuz, Zwaan, & Graesser, 1995; Shah, Evens, Michael, & Rovick, 2002; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). This chapter discusses some of these analyses, particularly from the standpoint of metacognition, meta-communication, and meta-affect. A third approach is to manipulate the tutoring activities through trained human tutors or computer tutors and to observe the impact of the manipulations on learning gains (Chi et al., 2001, 2008; Graesser, Lu, et al., 2004; VanLehn et al., 2007). Manipulation studies allow us to infer what characteristics of the tutoring directly cause increases in learning gains, barring potential confounding variables.

Given that the focus of this handbook is on metacognition, it is not our goal to provide a comprehensive literature review of the anatomy of tutoring and to identify the factors that best explain the impact of tutoring on learning. There are already several papers, meta-analyses, and reviews that address that goal, as we have already cited. Instead, this chapter attempts to reconstruct what metacognitive, meta-communicative, and meta-affective knowledge exists in the minds of tutors and tutees. What do tutors know, for example, about cognition, emotions, pedagogy, discourse, and the nature of tutoring? What do the tutees (students) know? Does such “meta” knowledge vary with the expertise of the tutor or tutee?

Most of what we know about the “meta” knowledge of tutoring is inferred from the tutoring process and available outcome measures. Traditional studies of metacognition outside of the tutoring realm have collected ratings or judgments from participants on their “meta” knowledge, such as feeling of knowing (FOK), judgments of learning (JOL), comprehension calibration, or inventories on the use of strategies (Azevedo & Cromley, 2004; Dunlosky, Rawson, & Middleton, 2005; Hacker, 1998). Such explicit judgments are not routinely collected in tutoring studies. For example, researchers do not collect data at different points in the tutoring session on whether the tutor knows or believes a fact *F*, the student knows *F*, the tutor knows the student knows *F*, or the student knows that the tutor knows that the student knows *F*. Researchers do not collect data, at different points in the tutoring session, on whether pedagogical strategy *S* should be used, or not used, and why. Researchers have not collected data on the general expertise that particular tutors have about good tutoring practices, the nature of cognitive processes, and the conditions in which a particular strategy would help learning. Such inventories would be useful to collect, but we were not able to identify a study that reported such data. The data we have available about tutoring is restricted to outcome measures of learning and/or the measures collected on the tutoring activities by trained researchers who analyze sessions recorded on videotape, audiotape, or computer (e.g., transcripts of computer-mediated interaction).

Nevertheless, much can be learned from tutoring process data and outcome measures. Consider a couple of examples that illustrate what we can infer. One of the hallmarks of tutoring is that the tutor should help the tutee correct deficits in his or her knowledge and skills. This is a large part of what makes tutoring adaptive. If a tutor knows this, the tutor should initiate some activity at the beginning of the session that diagnoses what the tutee does or does not know, or what challenges the tutee is struggling with. We can infer that

the tutor has this principle of adaptation if the tutor performs any of a number of actions: (a) inspects previous test materials and scores of the tutee, (b) selects problems in the tutoring session that are associated with the tutee's deficits, or (c) asks the tutee what they are having problems with. We can infer that the tutor lacks the principle of adaptation if the tutor immediately launches into problems from a curriculum script, in the same way for all tutees. This example illustrates that we can infer quite a bit by systematically analyzing the actions that occur during tutoring and comparing these actions to theoretically expected actions.

Consider another example that addresses metacognitive knowledge of the tutee. Tutors frequently ask students comprehension-gauging questions, such as "Do you understand?", "Are you following?", and "Does that make sense?" If the tutee's comprehension calibration skills are accurate, then the tutee should answer YES when the tutee understands and NO when there is little or no understanding. One provocative finding in the tutoring literature is that there sometimes is a positive correlation between a student's knowledge of the material (based on pre-test or post-test scores) and their likelihood of saying NO rather than YES to the tutors' comprehension-gauging questions (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Graesser et al., 1995). So it is the knowledgeable tutees who tend to say "No, I don't understand." This result suggests that deeper learners have higher standards of comprehension (Baker, 1985; Otero & Graesser, 2001) and that many students have poor comprehension calibration skills. The finding that students have sub-par comprehension calibration is well documented in the metacognitive literature, where meta-analyses have shown only a 0.27 correlation between comprehension scores on expository texts and the students' judgments on how well they understand the texts (Dunlosky & Lipko, 2007; Glenberg, Wilkinson, & Epstein, 1982; Maki, 1998). In this example, we can infer something about the tutee's metacognition by comparing a discourse pattern between tutor-tutee and a test score (pre-test or post-test).

The remainder of this chapter has two sections. We first discuss what tutees know about metacognition, meta-communication, and meta-affect on the basis of activities manifested in tutoring. As a note, we have a broad umbrella of what is considered "meta"; it is defined as their knowledge of pedagogy, cognition, communication, and emotion. After we cover what is known about the meta-knowledge of tutees we will turn to that of tutors. The tutors will include untrained human tutors, expert tutors, and computer tutors. Throughout this chapter, we identify multiple illusions that arise from metacognitive dysfunction that can ultimately undermine the success of tutored learning.

Meta-Knowledge of Tutees

Consider a sketch of what we would want from ideal learners. They would self-regulate their learning by identifying their own knowledge deficits, detecting contradictions, asking good questions that rectify these anomalies, searching knowledge sources for answers, making inferences when answers are not directly available, and actively building knowledge at deep levels of mastery. It is safe to say that the vast majority of tutees do not fit this profile (Chi, Siler, & Jeong, 2004; Graesser & Person, 1994; Graesser, McNamara, & VanLehn, 2005). It is the tutor who directs the tutoring session, not the tutee, and tutees are rarely inquisitive question askers (Chi et al., 2001; Graesser et al., 1995; Otero, this volume).

The more accurate sketch of the tutee's conception of a tutoring session is that of a passive learner, not too different from a student in a classroom listening to a lecture. They expect the tutor to lecture, explain the content, model procedures in solving problems, and essentially run the session. The tutee also expects to be tested with questions, just as in

the classroom. As in a classroom, they expect to be grilled with simple questions that invite short answers and clear-cut feedback on the quality of the answers. The Initiate-Respond-Evaluate (IRE) sequence in a classroom consists of the teacher initiating a question, the student giving a short-answer response, and the teacher giving a positive or negative evaluation of the response (Mehan, 1979; Sinclair & Coulthart, 1975). The analogue in the tutoring session would be the exchange below on the subject matter of Newtonian physics.

- (1) TUTOR: According to Newton's second law, force equals mass times what?
- (2) STUDENT: acceleration
- (3) TUTOR: Right, mass times acceleration.

Or

- (2) STUDENT: velocity
- (3) TUTOR: Wrong, it's not velocity, it is acceleration.

It is not surprising that the tutee's conception of a tutoring session is not much different from a classroom because the dynamics of the tutorial session closely mirror the classroom that they experience every day. Classrooms typically consist of the teacher presenting didactic lessons aligned with a curriculum, of presenting problems and worked-out solutions, and of frequently interrogating the students with IRE sequences. As in the classroom, the tutee expects to be evaluated and graded on their responses. Obviously, there are more innovative classroom environments that deviate from this simple sketch, but this does depict most classrooms.

Shortly after the tutee arrives at a tutoring session the tutee learns that the discourse and pedagogical structure of a tutoring session is somewhat different from the typical classroom. Although there is a tendency for poor tutors to simply lecture like a teacher, most tutors spend considerable time presenting problems or asking difficult questions that are answered collaboratively by the tutor and tutee (Chi et al., 2001; Graesser et al., 1995; Person & Graesser, 1999). According to Graesser and Person (1994), tutors frequently implement the following 5-step tutoring frame:

- (1) TUTOR asks a difficult question or presents a problem.
- (2) STUDENT gives an initial answer.
- (3) TUTOR gives short feedback on the quality of the answer.
- (4) TUTOR and STUDENT have a multi-turn dialogue to improve the answer.
- (5) TUTOR assesses whether the student understands the correct answer.

It is quite apparent that this 5-step tutoring frame involves collaborative discussion, joint action, and pressure for the tutee to construct knowledge rather than merely receiving knowledge. The role of the tutor shifts from being a knowledge-teller to a guide for collaborative knowledge construction. The relevant metaphor shifts from a classroom to collaborative work.

Tutors exhibit considerable variation on how they run their tutoring sessions. Later in this chapter we will discuss what tutors do and don't do, as well as their meta-knowledge. From the standpoint of the tutee, the focus of the present section, the meta-knowledge appears to be quite limited when we examine available evidence. The remainder of this section discusses what we can reconstruct about the tutee's meta-knowledge of cognition, communication, and emotions.

Metacognition

Other chapters in this volume discuss at length the fact that children, college students, and adults are quite limited in their metacognitive knowledge and skills, as well as their ability to self-regulate their learning. The process of self-regulation of learning theoretically involves the learners constructing a plan, monitoring metacognitive activities, implementing learning strategies, and reflecting on their progress and achievements (Azevedo, 2005, this volume; Azevedo & Cromley, 2004; Pintrich, 2000; Winne, 2001; Zimmerman, 2001). Each of these phases can be decomposed further. For example, metacognitive monitoring can be decomposed into JOL, FOK, content evaluation, monitoring the adequacy of a strategy, and monitoring progress towards goals (Hacker, Dunlosky, & Graesser, 1998). Examples of learning strategies include searching for relevant information in a goal-directed fashion, taking notes, drawing tables or diagrams, re-reading, elaborating the material, making inferences, coordinating information sources such as text and diagrams, and summarizing content. Tutees are limited, just as most learners are limited, in knowing, mastering, and executing most of these components.

Tutors are often struck by the lack of curiosity and questions of tutees. The lack of student questions can partially be attributed to the fact that the tutor governs the tutoring agenda. However, another reason is that there is a low rate of student questions in classrooms and most other learning environments (Dillon, 1988; Graesser, McNamara, & VanLehn, 2005). According to Graesser and Person's (1994) analysis of student questions, a typical student in a classroom asks approximately 1 question every 6–7 hours in a classroom environment and 1 question every 2 minutes in a tutoring environment. Most of the tutee questions are attempts to confirm or verify their knowledge (e.g., "Isn't this Newton's second law?"), to request definitions of terms ("What is a Newton?"), or to ask how to perform a procedure ("How do you get this square root?"), as opposed to questions that attempt to fill major gaps or contradictions in conceptual knowledge. Students with more domain knowledge ask deeper questions (Graesser & Olde, 2003; Graesser & Person, 1994; Miyake & Norman, 1979; Otero, this volume; Van der Meij, 1990). As Miyake and Norman (1979) pointed out decades ago in their analyses of tutoring, it takes a considerable amount of knowledge for a student to know what he or she does not know. According to one model of student question asking (Graesser & Olde, 2003; Otero & Graesser, 2001), students need to be put in cognitive disequilibrium before they ask sincere information-seeking questions. Cognitive disequilibrium occurs when students are confronted with contradictions, anomalies, obstacles to goals, and difficult decisions among equally attractive options. These conditions rarely occur when students are in a passive learning environment, they have minimal prior knowledge, and no vested interest in the subject matter. From this perspective, it is no surprise that good, deep, sincere information-seeking questions are rarely asked by the tutee.

As will be discussed later, tutors who adopt a student-centered perspective (Chi et al., 2001, 2008) attempt to get the tutees to do the talking and doing. The tutors generate pumps, prompts, hints, and questions to stimulate verbal explanations and actions of the tutee. Other tutors adopt a tutor-centered approach that is comparatively insensitive to the student's knowledge and the importance of the student constructing the knowledge. It is remarkable how few students, tutors, and members of the community fail to appreciate the value of active construction of knowledge. Some students may even feel cheated when they are expected to do the work rather than having the tutor do the knowledge-telling or model the actions. And the same holds for the parents. Parents occasionally criticize same-age peer tutoring because they strongly believe that teachers and tutors should do the teaching rather than having their children wasting their time

teaching other students. These parents do not understand the virtues of active learning and learning by teaching. At this point in tutoring research, it is an open question precisely what meta-knowledge students (and parents) have about the cognitive and pedagogical mechanisms of tutoring.

Metacommunication

Metacommunication is knowledge about the process of communication. The focus of this chapter is on tutoring and the focus of this section is on the tutee, so the relevant question addresses what knowledge the students have about the communication process during tutoring. Once again, available relevant data relies on an analysis of discourse patterns in conjunction with measures of student ability or achievement.

Communication is most successful when there is an alignment at multiple levels of language and discourse between participants (Pickering & Garrod, 2004). This occurs when there is high common ground (shared knowledge) and the pragmatic ground rules are well understood by both parties (Clark, 1996; Schober & Clark, 1989). Alignment violations sometimes occur during tutoring, however, because there is a large gap between tutor and tutee in subject matter knowledge and because the tutee sometimes misunderstands the tutor's tutoring goals (Person et al., 1995). This is most apparent when we examine the grounding of referents, feedback, and hints. A good tutor presumably attempts to clarify the conversational ground rules and to repair misalignments, but this too often is not achieved in normal tutoring sessions.

Grounding Referents

A referent is grounded when both the tutor and the student know the meaning of the referent and this meaning is shared in the discourse space (i.e., the history of the multi-turn dialogue and the learning materials that are co-present). Sometimes the meaning consists of the definition of a concept, but more often it is an entity in the tutoring environment being referred to or pointed to (e.g., a sentence in a book, a symbol in a math problem). Unfortunately, many students do not appreciate the importance of grounding so they let too much go by with a minimal understanding. When the tutor asks whether the tutee understands via a comprehension-gauging question ("Do you understand?"), they usually nod or answer YES, even though many of the referents are ungrounded and their understanding is minimal. As mentioned earlier, it is the more knowledgeable tutee who tends to answer NO (Chi et al., 1989; Graesser et al., 1995). Many tutors believe the student's answer and assume that conversational grounding has been achieved, so there is a serious misalignment in grounding. A good tutor might periodically show scepticism and make the student demonstrate adequate grounding by requesting the student to define a term, to point to an external referent, or to perform a procedure; that would be quite diagnostic of breakdowns in grounding. On the flip side, students often assume that the tutor understands whatever the student expresses. In essence, the student assumes "If I say it, the tutor will understand it." The truth is, however, that tutors often misunderstand the tutee because much of what the student says is vague, underspecified, and error-ridden (Graesser et al., 1995). The tutor's and tutee's illusion of grounding is pervasive in tutoring sessions and accounts for many of the misunderstandings and misalignments of communication. Table 19.1 defines this illusion, as well as other illusions that will be discussed throughout this chapter.

Table 19.1 Frequent Illusions during the Tutoring Process

Illusion of grounding. The unwarranted assumption that the speaker and listener have shared knowledge about a word, referent, or idea being discussed in the discourse context.

Illusion of feedback accuracy. The unwarranted assumption that the feedback that the other person gives to a speaker's contribution is accurate.

Illusion of discourse alignment. The unwarranted assumption that the listener does or is expected to understand the discourse function, intention, and meaning of the speaker's dialogue contribution.

Illusion of student mastery. The unwarranted assumption that the student has mastered much more than the student has really mastered.

Illusion of knowledge transfer. The speaker's unwarranted assumption that the listener understands whatever the speaker says and thereby knowledge is accurately transferred.

Feedback

Tutors give short feedback after most student turns as an indication of the quality of the student's contribution. The feedback is either positive (yes, very good, head nod, smile), negative (no, not quite, head shake, pregnant pause, frown), or neutral (uh huh, I see). After the short feedback, the tutor goes on to advance the conversation with questions, hints, elaborations, and other types of dialogue moves. The tutor's short feedback is hopefully accurate because many researchers believe that feedback is an important part of learning, although evidence for the impact of feedback on learning is mixed (Azevedo & Bernard, 1995; Chi et al., 2008; Kluger & DiNisi, 1998; Kulhavy & Stock, 1989; Shute, 2007). On the other hand, many tutors tend to be polite conversation partners and resist giving negative feedback to the tutee when the student expresses vague or error-ridden contributions (Person et al., 1995). Graesser and Person (1994) reported that the immediate feedback of unskilled tutors tends to have a higher likelihood of being positive than negative after incorrect or vague student contributions (Graesser et al., 1995). One could imagine that the students might clam up when they are barraged with a hefty dose of negative feedback, so there indeed is some virtue in tutors following the politeness principle. If tutor feedback does engender a trade-off between learning and motivation, then tutors should be particularly sensitive to the use of feedback. At the very least, they should be vigilant of the illusion of feedback accuracy.

What do students believe and expect regarding feedback? There is some anecdotal evidence suggesting that students do believe the feedback that they receive from tutors and also that they desire accurate feedback. For example, a computer tutor called AutoTutor helps students learn by holding a conversation with the student in natural language (Graesser, Lu, et al., 2004; Graesser, Person, Harter, & TRG, 2001). AutoTutor sometimes does not accurately interpret the tutee's contribution, which leads to incorrect short feedback. Students who detect the incorrect feedback get annoyed, irritated, or angry, sometimes to the point of dismissing the utility of AutoTutor altogether (D'Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008). One explanation of their negative emotions is that the students expect accurate feedback. Another observation from these sessions with AutoTutor is that students want decisive feedback (yes versus no) rather than evasive or indecisive feedback (possibly, uh huh, okay). A polite or wishy-washy computer tutor does not seem to be as desirable as a decisive one. It appears that the politeness principle may apply to humans but not computers in the case of feedback. These findings suggest that the students' assumptions about pragmatic ground rules and communication may be very different for human tutors versus computer tutors.

Hints

Good tutors are known to give hints that invite students to do the talking and doing rather than just giving lectures and telling the student the correct information (Chi et al., 2004, 2008; DiPaolo, Graesser, White, & Hacker, 2004; Hume, Michahael, Rovick, & Evens, 1993). The hints vary from being generic statements or questions (What about X?, Why?) to speech acts that more directly lead the student to a particular answer. Hints serve as memory cues, problem-solving clues, and directives to shift the burden of conversation from the tutor to the tutee. Hints may be the ideal scaffolding move to promote active student learning, but at the same time directing the student to focus on important relevant material.

Two limitations of hints potentially arise from the standpoint of student perceptions. Both of these limitations can be explained by appealing to an illusion of discourse alignment, the unwarranted assumption that the listener does or is expected to understand the discourse function, intention, and meaning of the speaker's dialogue contributions. The first limitation is that the students may not understand the discourse function of hints, particularly if they believe the ground rules dictate that the tutor should do the telling and showing. Students may have a tutor-centered epistemology of the tutoring register (i.e., genre) rather than a student-centered or interaction-centered epistemology (Chi et al., 2004, 2008). As a consequence, the student may not know how to respond to hints because they regard them as confusing or pragmatically infelicitous. A second limitation occurs when (a) the tutors have an answer in mind when they give a hint, (b) the students respond with an answer they think is good, (c) the students' responses do not match the tutors' expected answers, (d) the tutors give negative feedback, and (e) the students get frustrated or confused from being misunderstood and not adequately credited. D'Mello et al. (2008) have indeed reported that student confusion tends to occur after tutor hints and after negative feedback to students' responses to hints. It is interesting to speculate how tutorial dialogues would evolve if students assumed that many contributions in the conversation have discourse misalignments and that perfectly meshed conversations are very rare.

Meta-Affect

Connections between complex learning and emotions have received increasing attention in the fields of psychology (Carver, 2004; Deci & Ryan, 2002; Dweck, 2002) and education (Gee, 2003; Lepper & Henderlong, 2000; Linnenbrink & Pintrich, 2002; Meyer & Turner, 2002, 2006). It is important to understand affect-learning connections in order to adequately train tutors and also to design engaging educational artifacts that range from responsive intelligent tutoring systems (de Vicente & Pain, 2002; Graesser, Person, Lu, Jeon, & McDaniel, 2005; Guhe, Gray, Schoelles, & Ji, 2004; Litman & Silliman, 2004) to entertaining media and games (Conati, 2002; Gee, 2003; Vorderer, 2003). Researchers in many different fields are familiar with Ekman's work on the six universal emotions (sadness, happiness, anger, fear, disgust, surprise) that are manifested through facial expressions (Ekman & Friesen, 1978) and paralinguistic features of speech (Scherer, 1986). However, these are not the emotions that occur during most tutoring sessions. The pervasive affective states during the tutoring of technical material are confusion, frustration, boredom, anxiety, and flow/engagement, with delight and surprise occurring less frequently (Craig, Graesser, Sullins, & Gholson, 2004; D'Mello, Craig, Sullins, & Graesser, 2006; D'Mello, Picard, & Graesser, 2007; D'Mello et al., 2008; Lehman, Matthews, D'Mello, & Person, 2008).

Meyer and Turner (2006) identified three theories that are particularly relevant to understanding the links between emotions and learning: academic risk-taking, flow, and

goals. The academic risk theory contrasts (a) the adventuresome learners who want to be challenged with difficult tasks, take risks of failure, and manage negative emotions when they occur and (b) the cautious learners who tackle easier tasks, take fewer risks, and minimize failure and the resulting negative emotions (Clifford, 1991). According to flow theory, the learner is in a state of flow (Csikszentmihali, 1990) when the learner is so deeply engaged in learning the material that time and fatigue disappear. When students are in the flow state, they are at an optimal zone of facing challenges and conquering them by applying their knowledge and skills. Goal theory emphasizes the role of goals in predicting and regulating emotions (Dweck, 2002; Stein & Hernandez, 2007). Outcomes that achieve challenging goals result in positive emotions whereas outcomes that jeopardize goal accomplishment result in negative emotions.

Obstacles to goals are particularly diagnostic of both learning and emotions. The affective state of confusion correlates with learning gains perhaps because it is a direct reflection of deep thinking (Craig et al., 2004; Graesser, Chipman, King, McDaniel, & D'Mello, 2007). Confusion is diagnostic of cognitive disequilibrium, a state that occurs when learners face obstacles to goals, contradictions, incongruities, anomalies, uncertainty, and salient contrasts (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; Otero & Graesser, 2001). Cognitive equilibrium is ideally restored after thought, reflection, problem solving and other effortful deliberations. It is important to differentiate being productively confused, which leads to learning and ultimately positive emotions, from being hopelessly confused, which has no pedagogical value.

There have been very few studies on meta-affect in tutees. This leaves the door wide open for some informed speculation and some encouragement to researchers to conduct more research in this area. We do know that tutees identify the following affective states when they are asked to emote aloud (i.e., articulate their emotions) during learning or when they view videotapes of their tutoring sessions and judge their emotions at different points in time (D'Mello et al., 2006; Graesser et al., 2006): confusion, frustration, boredom, flow/engagement, delight, and surprise. However, we do not know how reliably different classes of learners can identify these emotions. We suspect from 150 years of psychological research on emotions that some learners lack sensitivity to their own emotions, that other learners are hypersensitive, and that there is a large continuum of possibilities in between (see Lewis, Haviland-Jones, & Barrett, 2008).

Research is conspicuously absent on how the tutees perceive the causes and consequences of these emotions and what they think they should do to regulate each affect state. The negative emotions are particularly in need of research. When a tutee is frustrated from being stuck, the student might attribute the frustration either to themselves ("I'm not at all good at physics"), the tutor ("My tutor doesn't understand this either"), or the materials ("This must be a lousy textbook"). Solutions to handle the frustration would presumably depend on these attributions of cause (Heider, 1958; Weiner, 1995). When a student is confused, some students may view this as a positive event to stimulate thinking and show their metal in conquering the challenge; other students will attribute the confusion to their poor ability, an inadequate tutor, or poorly prepared academic materials. When a student is bored, they are likely to blame the tutor or material rather than themselves. Tutoring environments of the future need to manage the tutorial interaction in a fashion that is sensitive to the tutees' emotions in addition to their cognitive states.

Meta-Knowledge of Tutors

What might we expect of the tutors' knowledge about pedagogy, metacognition, meta-communication, and meta-affect? In an ideal world, the tutor would be able to reliably

identify the cognitive and emotional states of the tutee. The tutor would be able to respond to these states in a pedagogically justifiable manner. That would include a large repertoire of excellent tutoring and teaching strategies that are tuned to different classes of students. The tutor would understand communication processes and how to formulate discourse moves that establish common ground and that advance the conversation in a manner that optimizes pedagogical goals.

The typical tutors in the school system and life-long learning are far from ideal, as we expressed earlier in the chapter. It is rare to have an experienced tutor with adequate training in the subject matter knowledge, pedagogy, and discourse processes. The review of studies with accomplished tutors by Person et al. (2007) makes the points that the sample sizes of expert tutors are extremely small ($N < 3$) in empirical investigations of expert tutoring, that the same expert tutors are used in different research studies, and the tutors are frequently co-authors on publications. Therefore, claims about expert tutoring are frequently biased by the idiosyncratic characteristics of the small sample and the tutors' authorship role. Person et al. (2007) recently conducted a study on a sample of 12 tutors who were nominated by teachers in the Memphis community who were truly outstanding. The discourse patterns of these outstanding tutors in Person's *Expert Tutor* corpus were dissected in great detail, but not linked to outcome scores on learning. The discourse patterns of 13 normal unskilled tutors, i.e., cross-aged tutors and paraprofessional, have been analyzed by Graesser and Person (Graesser & Person, 1994; Graesser et al., 1995; Person & Graesser, 1999, 2003). This Graesser–Person *Unskilled Tutor* corpus also was analyzed in great detail at the level of discourse, but not at the level of outcome scores on learning. Future research will need to systematically relate the discourse characteristics of tutors of varying ability to different outcome measures.

This section will sometimes consider *Computer Tutors* in addition to the Expert and Unskilled Tutors. The advantage of computer tutors is that they systematically analyze the students' knowledge and actions and they implement strategies to adaptively respond to the student. Unlike human tutors, the tutoring strategies are implemented systematically and follow complex algorithms.

Pedagogy

Unskilled human tutors are not particularly sophisticated from the standpoint of ideal tutoring strategies that have been proposed in the fields of education and artificial intelligence (Graesser et al., 1995). Graesser and colleagues videotaped over 100 hours of naturalistic tutoring in the Unskilled Tutoring corpus, transcribed the data, classified the speech act utterances into discourse categories, and analyzed the rate of particular discourse patterns. These analyses revealed that the tutors rarely implement intelligent pedagogical techniques such as *bona fide* Socratic tutoring strategies, modeling-scaffolding-fading, reciprocal teaching, frontier learning, building on prerequisites, or diagnosis/remediation of deep misconceptions (Collins, Brown, & Newman, 1989; Palincsar & Brown, 1984; Sleeman & Brown, 1982). In Socratic tutoring, the tutor asks students illuminating questions that lead the learners to discover and correct their own misconceptions in an active, self-regulated fashion (Collins, Warnock, Aeillo, & Miller, 1975). In modeling-scaffolding-fading, the tutor first models a desired skill, then gets the learners to perform the skill while the tutor provides feedback and explanation, and finally fades from the process until the learners perform the skill by themselves (Rogoff & Gardner, 1984). In reciprocal teaching, the tutor and learner take turns working on problems or performing a skill, as well as giving feedback to each other along the way (Palincsar & Brown, 1984). Tutors who use frontier learning select problems and give guidance in a fashion that

slightly extends the boundaries of what the learner already knows or has mastered. Tutors who build on prerequisites cover the prerequisite concepts or skills in a session before moving to more complex problems and tasks that require mastery of the prerequisites (Gagne, 1985).

It is perhaps unsurprising that unskilled human tutors rarely implement sophisticated tutoring strategies such as the ones described above. They were never trained to implement such strategies and it is unlikely these sophisticated strategies will be spontaneously discovered by the tutor. Moreover, there is a large computational overhead to implementing these strategies, far beyond most of the computer technologies of today. Researchers in computational linguistics and artificial intelligence have designed computer tutors to simulate tutors who interact with the students in natural language. These include AutoTutor (Graesser et al., 1999, 2001, 2004, 2006; VanLehn et al., 2007), why-Atlas (Graesser, VanLehn, Rose, Jordon, & Harter, 2001; Van-Lehn et al., 2003), CIRCSIM-Tutor (Shah et al., 2002), DC-Trains (Peters, Bratt, Clark, Pon-Barry, & Schultz, 2004), and Mission Rehearsal (Gratch et al., 2002). These different computer tutors vary in the extent to which they simulate human dialogue mechanisms, but all of them attempt to comprehend natural language and to formulate adaptive responses and strategies to help students learn.

The structure of the dialogue in both AutoTutor and human tutoring (Chi et al., 2001, 2004, 2008; Graesser, Hu, & McNamara, 2005; Shah et al., 2002) follows an Expectation and Misconception Tailored (EMT) dialogue. EMT dialogue is the primary pedagogical method of scaffolding good student answers. Both AutoTutor and human tutors typically have a list of expectations (anticipated good answers) and a list of anticipated misconceptions associated with each main question. For example, expectations E1 and E2 and misconceptions M1 and M2 are relevant to the example physics problem below:

PHYSICS QUESTION: If a lightweight car and a massive truck have a head-on collision, upon which vehicle is the impact force greater? Which vehicle undergoes the greater change in its motion, and why?

E1. The magnitudes of the forces exerted by A and B on each other are equal.

E2. If A exerts a force on B, then B exerts a force on A in the opposite direction.

M1: A lighter/smaller object exerts no force on a heavier/larger object.

M2: Heavier objects accelerate faster for the same force than lighter objects.

AutoTutor guides the student in articulating the expectations through a number of dialogue moves: pumps (what else?), hints, and prompts for the student to fill in missing words. Hints and prompts are carefully selected by AutoTutor to produce content in the answers that fill in missing content words, phrases, and propositions. For example, a hint to get the student to articulate expectation E1 might be “What about the forces exerted by the vehicles on each other?”; this hint would ideally elicit the answer “The magnitudes of the forces are equal.” A prompt to get the student to say “equal” would be “What are the magnitudes of the forces of the two vehicles on each other?” As the learner expresses information over many turns, the list of expectations is eventually covered and the main question is scored as answered. Complete coverage of the answer requires AutoTutor to have a pool of hints and prompts available to extract all of the content words, phrases, and propositions in each expectation. AutoTutor adaptively selects those hints and prompts that fill missing constituents and thereby achieves pattern completion.

Human tutors and AutoTutor are dynamically adaptive to the learner in ways other than coaching them to articulate expectations. There is the conversational goal of correcting misconceptions that arise in the student’s responses. When the student articulates a misconception, the tutor acknowledges the error and corrects it. There is another

conversational goal of giving feedback to the student on their contributions. For example, the tutor gives short feedback on the quality of student contributions, as discussed earlier. The tutor accommodates a mixed-initiative dialogue by attempting to answer the student's questions when the student is sufficiently inquisitive to ask questions. The tutor asks counter-clarification questions (e.g., "I don't understand your question, so could you ask it in another way?") when the tutor does not understand the student's question. Tutors are considered more adaptive to the student to the extent that they correct student misconceptions, give correct feedback, answer student questions, and ask clarification questions to ensure the grounding of content. AutoTutor and other dialogue-based intelligent tutoring systems implement these features of conversational responsiveness.

In addition to engaging in EMT dialogue, unskilled human tutors and AutoTutor have a case-based learning foundation. That is, they present challenging problems for the student to collaboratively reason with the tutor in route to an answer. The hope is that deep learning eventually emerges after the student and tutor collaboratively attempt to solve a large number of problems that vary in scope and sophistication. However, the cases are not necessarily anchored in real-life situations (Bransford, Brown, & Cocking, 2000) because those problems require a lengthy process of science and engineering to design.

The general message to be conveyed here is that sophisticated pedagogical strategies are not generated by unskilled tutors, which is perhaps unsurprising because these strategies are complex and took centuries to discover by scholars. However, it is a very important finding to document because it is conceivable that deep learning could improve tremendously by training human tutors and programming computer tutors to implement the sophisticated strategies. It is a question for future research to determine whether it is technically feasible for these strategies to be reliably implemented in computers and trained human tutors. It is also an open question whether the strategies significantly improve learning over and above the normal strategies of unskilled human tutors.

The pedagogical strategies of expert tutors are very similar to those of unskilled tutors and computer tutors in most ways. However, Person et al. (2007) identified a few improvements in the pedagogy in the Expert Tutor corpus. The expert tutors did occasionally implement modeling-scaffolding-fading, although the relative frequencies of the dialogue moves for this pedagogical strategy were not high. One pedagogical strategy evident in the Expert Tutor corpus was just-in-time direct instruction or mini-lectures when the tutee was struggling with a particular conceptualization. These content-sensitive lectures were sensitive to what the tutee was having trouble with rather than being routinely delivered to all tutees. The expert tutors also appeared to differ from unskilled tutors on metacognitive and meta-communicative dimensions, to which we now turn.

Metacognition

The essential question is whether the tutor has knowledge of the cognitive states of the tutee. The tutor's knowledge of the tutor's own cognitive states is not directly relevant to this section; presumably the metacognitive proficiencies of tutors are equal or occasionally higher than the tutees, particularly in the case of expert tutors. We would hope that the tutor builds an accurate and detailed model of the cognitive states of the tutee, or what is called the student model by researchers who develop intelligent tutoring systems. Thus, the central question is how accurate and detailed the student model is that gets constructed and used by the tutor.

There are reasons for being pessimistic about the quality of the student model that tutors construct. A more realistic picture is that the tutor has only an approximate appraisal of the cognitive states of tutees and that they formulate responses that do not require fine

tuning of the student model (Chi et al., 2004; Graesser et al., 1995). What is the evidence for such claims? There are three sources of evidence. First, the short feedback to students on the quality of the students' contributions is often incorrect. For example, the short feedback has a higher likelihood of being positive than negative after student contributions that are vague or error-ridden (Graesser et al., 1995). Second, tutors do not have a high likelihood of detecting misconceptions and error-ridden contributions of students (Chi et al., 2004; VanLehn et al., 2007). Computer tutors also have a similar problem of having difficulties detecting errors in verbal descriptions (Graesser, Hu, & McNamara, 2005). Third, tutors do not select new cases or problems to work on that are sensitive to the abilities and knowledge deficits of students (Chi et al., 2008). You would expect the selection of problems to be tailored to the tutee's profile according to the Goldilock's principle or zone of proximal development, i.e., not too easy or not too hard, but just right. However, Chi et al. (2008) reported that there was no relation between problem selection and tutee's profile. Data such as these lead one to conclude that tutors have a modest ability to conduct student modeling. Perhaps the only student parameter that the tutor can reliably handle is the verbosity of the tutee, namely the extent to which they express their ideas, as measured in number of words. However, the depth of the tutee's contributions is apparently difficult to calibrate and the specific knowledge deficits extremely difficult to identify. It may take intelligent tutoring systems to detect such subtleties and execute context-sensitive adaptive responses to promote learning.

Tutors often fall prey to the illusion of student mastery, namely the belief that the tutee has mastered much more than they actually have. As discussed earlier in this chapter, tutors often believe the students when they ask the comprehension-gauging question "Do you understand?" and most students answer yes. The tutor often believes the tutee has covered an expectation, even when the students' contribution is vague or error-ridden. The tutors tend to miss misconceptions expressed by students. Tutors frequently believe that the student has expressed a complete sentence-like expectation, even when the student only articulates a couple of content words in the expectation. All of these findings point in the direction that the tutor gives the student the benefit of the doubt with respect to student modeling. They assume that the tutee has mastered whatever is said or covered in the tutorial session, even though (a) much of what is covered in the dialogue is vague, incomplete, and error-ridden and (b) the tutee does not understand much of what the tutor expresses.

It is conceivable that expert tutors are more successful in detecting and tracking the knowledge of the students (Person et al., 2007). This is manifested in the question-asking analyses that Person et al. performed on the Expert Tutor corpus and compared to the previously published studies on the Unskilled Tutor corpus (Graesser & Person, 1994; Person & Graesser, 1999). Expert tutors and students ask proportionately more low-specificity questions (e.g., So?) and more common ground questions (e.g., So, I use the Pythagorean theorem?) than tutors and students in non-expert sessions. We interpret these findings to mean that expert tutors are more attuned to the needs of their students and have established considerable common ground. If this were not the case, low-specificity questions (e.g., So?) would result in conversation breakdowns. As another form of evidence, expert tutors are more dynamic in their instruction and do not rely on curriculum scripts. Experts typically begin the tutoring sessions by figuring out the topics/problems that students are having difficulty with and by asking questions about the students' performance on quizzes, homework, and exams. After this data collection phase, the tutor decides where to begin the session and what material will be covered. Expert tutors do not begin a session with any pre-planned teaching agenda, but rather base their instruction on students' particular needs at the time of the tutoring session.

One research question for the future addresses the fidelity of student modeling for unskilled, expert, and computer tutors. Which of these classes of tutors is most accurate in detecting the cognitive profile of the learners? How adaptive are the different tutors in producing discourse moves to facilitate learning in individual tutees? It is likely that the mental models of some students are so error-ridden and distorted that it would not be worth the time for the tutor to unravel the pathological conceptual spaghetti. When this occurs, the best the tutor can do is to model good reasoning and strategies. However, some students may have mental models that approximate some semblance of accuracy. These are the students who might benefit from accurate student modeling and intelligent tutor responses. Answers to these questions await further research.

Meta-Communication

Speakers typically assume that the meaning of their messages is correctly transferred to the minds of the addressees. In the tutoring context, knowledge transfer is successful when the tutee understands whatever the tutor expresses in the discourse space. However, there is abundant evidence, as we have discussed throughout this chapter, that tutees often do not understand the subject matter and the discourse contributions expressed by the tutor. The illusion of knowledge transfer occurs when the tutor assumes that the student understands whatever the tutor says. A good tutor is aware of this potential illusion. A good tutor is sufficiently skeptical of the tutee’s level of understanding that the tutor troubleshoots potential communication breakdowns between the tutor and tutee. The tutor does this by not trusting the tutee’s answer to comprehension gauging questions (“Do you understand?”) and instead asking follow-up questions that verify the student’s understanding. This is illustrated in the simple exchange below:

- Tutor:* We know from Newton’s law that net force equals mass times acceleration.
This law. . . .
- Student:* Yeah, that is Newton’s second law.
- Tutor:* Do you get this?
- Student:* Yeah. I know that one.
- Tutor:* Okay, let’s make sure. Force equals mass times what?
- Student:* times velocity.
- Tutor:* No, it’s mass times acceleration.

Students frequently have low domain knowledge so they do not reliably attend to, understand, or remember what the tutor says. They have difficulty applying abstract principles to concrete applications. Therefore, a good tutor should not fall prey to the illusion of knowledge transfer. A good tutor assumes that the student understands very little of what the tutor says and that knowledge transfer approaches zero. Person et al. (2007) reported that expert tutors are more likely to verify that the tutee understands what the tutor expresses by asking follow-up questions or giving follow-up troubleshooting problems.

Knowledge transfer is facilitated when the tutor gives accurate feedback on the tutee’s contributions. As discussed earlier, unskilled tutors frequently give incorrect feedback, as in the case of expressing more positive than negative short feedback after vague or incorrect student contributions (Graesser et al., 1995). In contrast, Person et al. (2007) report that the expert tutors tend to give correct short feedback. Accountability of student understanding is most successful when the tutor questions have precise specific answers (as in mathematics) rather than a family of alternative imprecise answers (Person et al., 1995). Therefore, a good tutor should ask some discriminating questions with ideal answers that

are sufficiently clear-cut that they can quickly diagnose incorrect student understanding and misconceptions. Without such precise accountability of student knowledge, it is unlikely that much communication is taking place in a tutoring session.

Meta-Emotion

As discussed earlier, emotions play a critical role in the learning process. Therefore, it is presumably important for tutors to adopt pedagogical and motivational strategies that are effectively coordinated with the students' emotions (Issroff & del Soldato, 1996; Lepper & Chabay, 1988; Lepper & Woolverton, 2002). Lepper and Woolverton (2002) proposed an INSPIRE model to promote this integration. This model encourages the tutor to nurture the tutee by being empathetic and attentive to the tutee's needs, to assign tasks that are not too easy or difficult, to give indirect feedback on erroneous student contributions rather than harsh feedback, to encourage the tutee to work hard and face challenges, to empower the student with useful skills, and to pursue topics they are curious about. One of the interesting tutor strategies is to assign an easy problem to the tutee, but to claim that the problem is difficult and to encourage the student to give it a try anyway. When the tutee readily solves the problem, the student builds self-confidence and self-efficacy in conquering difficult material (Winne, 2001; Zimmerman, 2001).

A tutor that is aware of the student's affective states would be expected to adaptively respond to their emotions during the course of enhancing learning. For example, if the tutee is frustrated, the tutor might give hints to advance the tutee's construction of knowledge (Graesser, Rus, D'Mello, & Jackson, 2008) or make supportive empathetic comments to enhance motivation (Burlinson & Picard, 2004; Lepper & Woolverton, 2002). If the tutee is bored, a good tutor would present more engaging or challenging problems for the tutee to work on. The tutor would probably want to lay low and stay out of the student's way when the tutee is deeply engrossed in a state of flow (Csíkszentmihályi, 1990). The flow experience is believed to occur when the learning rate is high and the student has achieved a high level of mastery at the region of proximal learning (Metcalfe & Kornell, 2005). When the tutee is confused, there is a variety of productive paths for the tutor to pursue (D'Mello et al., 2007; Graesser et al., 2008). The tutor could allow the student to continue being confused during the cognitive disequilibrium; the student's self-regulated thoughts might hopefully restore equilibrium. Alternatively, after some period of time waiting for the student to progress, the tutor might give indirect hints to nudge the tutee into more productive trajectories of thought.

Goleman (1995) stated in his book, *Emotional Intelligence*, that expert teachers are able to recognize a student's emotional state and respond in an appropriate manner that has a positive impact on the learning process. Lepper and Woolverton (2002) have claimed that it takes expertise in tutoring before accurate detection of learner emotions can be achieved. This requirement of expertise is apparently quite important because, according to Lepper and Woolverton (2002), roughly half of expert tutors' interactions with the student are focused on affective elements.

These important claims would be seriously limited if tutors are unable to detect the affective states of the learner. Unfortunately, there is some evidence that humans untrained in detecting emotions are not particularly adept at detecting the emotions of a learner. For example, Graesser and colleagues conducted a study that assessed the reliability of emotion judgments of tutees, untrained peers, and accomplished teachers (D'Mello, Taylor, Davidson, & Graesser, 2008; Graesser et al., 2006). There were also two trained judges: research assistants who were trained extensively on tutorial dialogue characteristics and how to detect facial action units according to Paul Ekman's Facial Action Coding System

(Ekman & Friesen, 1978). The study involved 28 college students being tutored on computer literacy with AutoTutor. Judgments were made on the basis of videos of the participants' faces and screens that were recorded during the interaction.

Trained judges experienced in coding facial actions and tutorial dialogue provided affective judgments that were reliable and that matched the learner's self-reports better than the affect judgments of untrained peers and accomplished teachers. Training on facial expressions apparently makes judges more mindful of relevant facial features and transient facial movements. However, in many situations the face is not the best indicator of the emotions of the learner. For example, while confusion and delight are accompanied by animated facial expressions, the face is virtually expressionless during expressions of boredom and flow (McDaniel et al., 2007; Craig, D'Mello, Witherspoon, & Graesser, 2007). Therefore, experience with tutorial dialogue is also an important requirement for affect detection.

The finding that peers and teachers are not very accurate at classifying the emotions of the learner has considerable implications for peer tutoring and expert tutors. One potential advantage of peer tutoring is that there is no appreciable status difference between peers, compared to when a teacher tutors a student or when an older tutor helps a younger learner (Fuchs et al., 1994; Rogoff, 1990). There are several advantages of expert tutors who construct more accurate models of student knowledge, provide more evaluative and discriminating feedback, adapt their instructional strategies dynamically, and are more direct and task-oriented (Person et al., 2007). However, the results of the studies on emotion tracking (D'Mello et al., 2008; Graesser et al., 2006) suggest that the inability to detect the affective states of a learner might be a drawback of peer and teacher tutoring. We are uncertain at this point on the importance of meta-affective abilities, status differences, experience with domain knowledge, and pedagogy with respect to learning gains and emotion tracking. Future research is needed to resolve this.

Summary

This chapter has discussed the metacognitive, meta-communicative, and meta-affective mechanisms that occur in one-on-one human tutoring. We examined these capabilities from the perspective of both the tutor and the student. The typical unskilled tutors that exist in K-12 and college do not have a sophisticated repertoire of tutoring strategies that monitor the course of learning, such as building on pre-prerequisites, modeling-scaffolding-fading, Socratic tutoring, and misconception diagnosis and repair. These tutors are also limited in detecting the idiosyncratic cognitive and emotional profiles of individual students, generating dialogue moves that adapt to these profiles in a pedagogically sensitive manner, and incorporating most theoretical phases of metacognition and meta-affect. Instead, there is a strong tendency to follow curriculum scripts and dialogue tactics that monitor expected correct information and anticipated errors. Accomplished and trained tutors do exhibit a few enhancements in meta-knowledge, but even these are limited. From the perspective of the student, there are well-documented limitations in their ability to monitor and strategically augment their own knowledge and learning, whereas a handful of studies have examined how well they detect and manage their emotions. Given the limitations in the metacognitive, meta-communicative, and meta-affective capabilities in human tutoring, researchers are encouraged to increase their efforts in building trainers, intelligent tutoring systems, and other learning environments that attempt to improve these mechanisms.

Author Note

The research on tutoring was supported by the National Science Foundation (SBR 9720314, REC 0106965, REC 0126265, ITR 0325428, REESE 0633918), the Institute of Education Sciences (R305H050169, R305B070349, R305A080594), and the Department of Defense Multidisciplinary University Research Initiative (MURI) administered by ONR under grant N00014-00-1-0600. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, IES, or DoD. The Tutoring Research Group (TRG) is an interdisciplinary research team comprised of researchers from psychology, computer science, physics, and education (visit <http://www.autotutor.org>, <http://emotion.autotutor.org>, <http://fedex.memphis.edu/iis/>). Requests for reprints should be sent to Art Graesser, Department of Psychology, 202 Psychology Building, University of Memphis, Memphis, TN 38152-3230, a-graesser@memphis.edu

References

- Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist, 40*, 193–198.
- Azevedo, R. & Bernard, R. M. (1995). A meta-analysis of the effect of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*, 109–125.
- Azevedo, R. & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia. *Journal of Educational Psychology, 96*, 523–535.
- Baker, L. (1985). Differences in standards used by college students to evaluate their comprehension of expository prose. *Reading Research Quarterly, 20*, 298–313.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*, 4–16.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (2000). *How People Learn* (expanded ed.). Washington, DC: National Academy Press.
- Burleson, W. & Picard, R. W. (2004). Affective Agents: Sustaining Motivation to Learn Through Failure and a State of Stuck. *Social and Emotional Intelligence in Learning Environments Workshop In Conjunction with the 7th International Conference on Intelligent Tutoring Systems*. August 31, 2004. Maceio—Alagoas, Brasil.
- Carver, C. S. (2004). Negative affects deriving from the behavioral approach system. *Emotion, 4*, 3–22.
- Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182.
- Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutotial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science, 32*, 301–341.
- Chi, M. T. H., Siler, S. A., & Jeong, H. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction, 22*(3): 363–387.
- Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471–534.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clifford, M. M. (1991). Risk taking: Theoretical, empirical, and educational considerations. *Educational Psychologist, 26*, 263–298.
- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237–248.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Collins, A., Warnock, E. H., Aeillo, N., Miller, M. L. (1975). Reasoning from incomplete knowledge. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding* (pp. 453–494). New York: Academic.
- Conati C. (2002). Probabilistic assessment of user's emotions in educational games. *Journal Of Applied Artificial Intelligence*, 16, 555–575.
- Craig, S., D'Mello, S., Witherspoon, A., & Graesser, A. (2007). Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive-affective states during learning. *Cognition and Emotion*, 22, 777–788.
- Craig, S. D., Graesser, A. C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media*, 29, 241–250.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper-Row.
- Deci, E. L. & Ryan, R. M. (2002). The paradox of achievement: The harder you push, the worse it gets. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61–87). Orlando, FL: Academic Press.
- De Vicente, A. & Pain, H. (2002). Informing the detection of students' motivational state: An empirical study. *Intelligent Tutoring Systems 2002*, 933–943.
- Dillon, J. T. (1988). The remedial status of student questioning. *Journal of Curriculum Studies*, 20 (3), 197–210.
- DiPaolo, R. E., Graesser, A. C., White, H. A., & Hacker, D. J. (2004). Hints in human and computer tutoring. In M. Rabinowitz (Ed.), *The design of instruction and evaluation* (pp. 155–182). Mahwah, NJ: Erlbaum.
- D'Mello, S. K., Craig, S. D., Sullins, J., & Graesser, A. C. (2006). Predicting affective states through an emote-aloud procedure from Autotutor's mixed-initiative dialogue. *International Journal Of Artificial Intelligence In Education*, 16, 3–28.
- D'Mello, S. K., Craig, S. D., Witherspoon, A. W., McDaniel, B. T., & Graesser, A. C. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1–2), 45–80.
- D'Mello, S. K., Picard, R., & Graesser, A. C. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, 22, 53–61.
- D'Mello, S. K., Taylor, R. S., Davidson, K., & Graesser, A. C. (2008). Self versus teacher judgments of learner emotions during a tutoring session with AutoTutor. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent Tutoring Systems: 9th International Conference*. Heidelberg, Germany: Springer, pp. 9–18.
- Dunlosky, J. & Lipko, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, 16, 228–232.
- Dunlosky, J., Rawson, K. A., & Middleton, E. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language*, 52, 551–565.
- Dweck, C. S. (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 61–87). Orlando, FL: Academic Press.
- Ekman, P. & Friesen, W. V. (1978). *The facial action coding system: A technique for the measurement of facial movement*. Palo Alto: Consulting Psychologists Press.
- Fuchs, L., Fuchs, D., Bentz, J., Phillips, N., & Hamlett, C. (1994). The nature of students' interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, 31, 75–103.
- Gagne, R. M. (1985). *The conditions of learning and theory of instruction* (4th ed.). New York: Holt, Rinehart, & Winston.
- Gee, J. P. (2003). *What video games have to teach us about language and literacy*. New York: Macmillan.
- Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, 10, 597–602.
- Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.
- Graesser, A. C. & Olde, B. A. (2003). How does one know whether a person understands a

- device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, 95, 524–536.
- Graesser, A. C., Chipman, P., King, B., McDaniel, B., & D’Mello, S. (2007). Emotions and learning with AutoTutor. In R. Luckin, et al. (Eds.), *13th International Conference on Artificial Intelligence in Education (AIED 2007)* (pp. 569–571). IOS Press.
- Graesser, A. C., D’Mello, S. K., Craig, S. D., Witherspoon, A., Sullins, J., McDaniel, B., & Gholson, B. (2008). The relationship between affect states and dialogue patterns during interactions with AutoTutor. *Journal of Interactive Learning Research*, 19, 293–312.
- Graesser, A. C., Hu, X., & McNamara, D. S. (2005). Computerized learning environments that incorporate research in discourse psychology, cognitive science, and computational linguistics. In A. F. Healy (Ed.), *Experimental Cognitive Psychology and its Applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*. Washington, DC: American Psychological Association.
- Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, 36, 180–193.
- Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory and Cognition*, 33, 1235–1247.
- Graesser, A. C., McDaniel, B., Chipman, P., Witherspoon, A., D’Mello, S., & Gholson, B. (2006). Detection of emotions during learning with Autotutor. *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 285–290). Mahwah, NJ: Erlbaum.
- Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist*, 40, 225–234.
- Graesser, A. C. & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31, 104–137.
- Graesser, A. C., Person, N., Harter, D., & TRG (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, 12, 257–279.
- Graesser, A. C., Person, N., Lu, Z., Jeon, M., & McDaniel, B. (2005). Learning while holding a conversation with a computer. In L. PytlikZillig, M. Bodvarsson, & R. Bruning (Eds.), *Technology-based education: Bringing researchers and practitioners together* (pp. 143–168). Greenwich, CT: Information Age Publishing.
- Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, 9, 1–28.
- Graesser, A. C., Rus, V., D’Mello, S., & Jackson, G. T. (2008). AutoTutor: Learning through natural language dialogue that adapts to the cognitive and affective states of the learner. In D. H. Robinson & G. Schraw (Eds.), *Current perspectives on cognition, learning, and instruction* (pp. 95–126). Greenwich, CT: Information Age Publishing.
- Graesser, A. C., VanLehn, K., Rose, C., Jordan, R., & Harter, D. (2001). Intelligent Tutoring Systems with conversational dialogue. *AI Magazine*, Winter 2001, 39–51.
- Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the TRG (1999). Auto Tutor: A simulation of a human tutor. *Journal of Cognitive Systems Research*, 1, 35–51.
- Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., & Badler, N. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, 17, 54–63.
- Guhe, M., Gray, W. D., Schoelles, M. J., & Ji, Q. (2004, July). *Towards an affective cognitive architecture*. Poster session presented at the Cognitive Science Conference, Chicago, IL.
- Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1–24). Mahwah, NJ: Erlbaum.
- Hacker, D. J. & Graesser, A. C. (2007). The role of dialogue in reciprocal teaching and naturalistic tutoring. In R. Horowitz (Ed.), *Talk about text: How speech and writing interact in school learning* (pp. 253–270). Mahwah, NJ: Erlbaum.
- Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.) (1998). *Metacognition in educational theory and practice*. Mahwah, NJ: Erlbaum.

- Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley & Sons.
- Hume, G., Michael, J., Rovick, A., & Evens, M. (1993). The use of hints as a tutorial tactic. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 563–568). Boulder, CO.
- Issroff, K. & del Soldato, T. (1996). Incorporating motivation into computer-supported collaborative learning. In Brna, P., et al. (Eds.), *Proceedings of the European Conference on Artificial Intelligence in Education*. New York: John Wiley & Sons.
- Johnson, D. W. & Johnson, R. T. (1992). Implementing cooperative learning. *Contemporary Education*, 63(3), 173–180.
- King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology*, 90, 134–152.
- Kluger, A. N. & DiNisi, A. (1996) Feedback Interventions: Toward the understanding of a two-edged sword. *Current Directions in Psychological Science*, 7(3), 67–72.
- Kulhavy, R. W. & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1, 279–308.
- Lehman, B. A., Matthews, M., D'Mello, S. K., & Person, N. (2008). Understanding students' affective states during tutoring. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent Tutoring Systems: 9th International Conference* (pp. 50–59). Heidelberg, Germany: Springer.
- Lepper, M. R. & R. W. Chabay (1988). Socializing the intelligent tutor: Bringing empathy to computer tutors. In H. Mandl & A. Lesgold (Eds.), *Learning issues for Intelligent Tutoring Systems* (pp. 242–257). Hillsdale, NJ: Erlbaum.
- Lepper, M. R., Drake, M., & O'Donnell-Johnson, T. M. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues* (pp. 108–144). New York: Brookline Books.
- Lepper, M. R. & Henderlong, J. (2000). Turning “play” into “work” and “work” into “play”: 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 257–307). San Diego, CA: Academic Press.
- Lepper, M. R. & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135–158). Orlando, FL: Academic Press.
- Lewis, M., Haviland-Jones, J. M., & Barrett, L. F. (Eds.). (2008). *The handbook of emotion*, 3rd ed. New York: Guilford.
- Linnenbrink, E. A. & Pintrich, P. (2002). The role of motivational beliefs in conceptual change. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Litman, D. J. & Silliman, S. (2004). ITSPROKE: An intelligent tutoring spoken dialogue system. In *Proceedings of the human language technology conference: 3rd meeting of the North American chapter of the association of computational linguistics* (pp. 52–54). Edmonton, Canada: ACL.
- Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Mahwah, NJ: Erlbaum.
- Mathes, P. G. & Fuchs, L. S. (1994). Peer tutoring in reading for students with mild disabilities: A best evidence synthesis. *School Psychology Review*, 23, 59–80.
- McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction*, 7, 197–244.
- McDaniel, B. T., D'Mello, S. K., King, B. G., Chipman, P., Tapp, K., & Graesser, A. C. (2007). Facial features for Affective State Detection in learning environments. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 467–472). Austin, TX: Cognitive Science Society.
- Mehan, H. (1979). *Learning lessons: social organization in the classroom*. Cambridge, MA: Harvard University Press.
- Merrill, D. C., Reiser, B. J., Merrill, S. K., & Landes, S. (1995). Tutoring: Guided learning by doing. *Cognition and Instruction*, 13(3), 315–372.

- Metcalfe, J. & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, 52, 463–477.
- Meyer, D. K. & Turner, J. C. (2002). Discovering emotion in classroom motivation research. *Educational Psychologist*, 37(2), 107–114.
- Meyer, D. K. & Turner, J. C. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review*, 18 (4), 377–390.
- Miyake, N. & Norman, D. A. (1979). To ask a question one must know what is not known. *Journal of Verbal Learning and Verbal Behavior*, 18, 357–364.
- Otero, J. & Graesser, A. C. (2001). PREG: Elements of a model of question asking. *Cognition & Instruction*, 19, 143–175.
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension: Fostering and monitoring activities. *Cognition and Instruction*, 1, 117–175.
- Person, N. K. & Graesser, A. C. (1999). Evolution of discourse in cross-age tutoring. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 69–86). Mahwah, NJ: Erlbaum.
- Person, N. K. & Graesser, A. C. (2003). Pedagogical agents and tutors. In J. W. Guthrie (Ed.), *Encyclopedia of education* (2nd ed., Vol. 2, pp. 586–589). New York: Macmillan.
- Person, N. K., Kreuz, R. J., Zwaan, R., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction*, 13, 161–188.
- Person, N., Lehman, B., & Ozbun, R. (2007). Pedagogical and motivational dialogue moves used by expert tutors. Presented at the 17th Annual Meeting of the Society for Text and Discourse. Glasgow, Scotland.
- Peters, S., Bratt, E. O., Clark, B., Pon-Barry, H., Schultz, K. (2004). Intelligent Systems for Training Damage Control Assistants, in Proceedings of IITSEC 2004, Orlando FL, USA.
- Pickering, M. J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Brain and Behavioral Sciences*, 27, 169–190.
- Pilkington, R. & Parker-Jones, C. (1996) Interacting with computer based simulation: The role of dialogue. *Computers and Education*, 27, 1, 1–14.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452–502). New York: Academic Press.
- Rogoff, B. (1990). *Apprenticeship in thinking*. New York: Oxford University Press.
- Rogoff, B. & Gardner, W. (1984). Adult guidance of cognitive development. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 95–116). Cambridge, MA: Harvard University Press.
- Rohrbeck, C. A., Ginsburg-Block, M., Fantuzzo, J. W., & Miller, T. R. (2003). Peer assisted learning Interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, 95 (2), 240–257.
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99(2), 143–165.
- Schober, M. F. & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21, 211–232.
- Shah, F., Evens, M. W., Michael, J., & Rovick, A. (2002). Classifying student initiatives and tutor responses in human keyboard-to-keyboard tutoring sessions. *Discourse Processes*, 33, 23–52.
- Shute, V. J. (2007). *Focus on formative feedback* (ETS Research Rep. No. RR-07-11). Princeton, NJ: ETS.
- Sinclair, J. M. & Coulthard, R. M. (1975). *Towards an analysis of discourse: The English used by teachers and their pupils*. London: Oxford University Press.
- Slavin, R. E. (1990). *Cooperative learning: Theory, research, and practice*. New Jersey: Prentice Hall.
- Sleeman D. & J. S. Brown. (1982). Introduction: Intelligent Tutoring Systems. In *Intelligent Tutoring Systems* (pp. 1–10). Orlando, Florida: Academic Press, Inc.
- Stein, N. L. & Hernandez, M. W. (2007). Assessing understanding and appraisals during emotional

- experience: The development and use of the Narcoder. In J. A. Coan & J. J. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 298–317). New York: Oxford University Press.
- Topping, K. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, 32, 321–345.
- Van der Meij, H. (1990). Question asking: To know that you do not know is not enough. *Journal of Educational Psychology*, 82, 505–512.
- VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, 31, 3–62.
- VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3), 209–249.
- Vorderer, P. (2003). Entertainment theory. In B. Jennings, D. Roskos-Ewoldsen, & J. Cantor (Eds.), *Communication and emotion: Essays in honor of Dolf Zillmann* (pp. 131–153). Mahwah, NJ: Erlbaum.
- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford.
- Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Mahwah, NJ: Erlbaum.
- Zimmerman, B. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 1–37). Mahwah, NJ: Erlbaum.

20 *In Vivo* Experiments on Whether Supporting Metacognition in Intelligent Tutoring Systems Yields Robust Learning

*Ken Koedinger, Vincent Aleven, Ido Roll,
and Ryan Baker*
Carnegie Mellon University

Interactive educational technologies, like intelligent tutoring systems, provide an excellent platform on which to perform research on metacognition. Such systems provide fine-grained tracking and assessment of students' cognitive and metacognitive behaviors. They also facilitate the implementation of tightly controlled experiments within the context of real classrooms. This chapter covers recent research on the impact of different types of metacognitive support in intelligent tutoring systems on both the learning of domain content and desired metacognitive behaviors. We describe results from four lines of controlled experimentation, mostly performed in the context of real classrooms, *in vivo*. Our results show clear benefits of metacognitive supports for domain content learning. We also demonstrate innovative methods for assessing real-time metacognitive behaviors as students are engaging in learning activities. Finally, we present some partial success in attempts to produce lasting improvements in metacognitive behaviors themselves and discuss why metacognitive behaviors may be resistant to change.

Following Brown (1987), we define metacognition as thinking about cognition (memory, perception, reasoning, etc.) itself, that is, reasoning about one's own thinking. Within this chapter, we follow Schoenfeld's (1987) framing of "metacognitive learning strategies" as specific kinds/uses of metacognition that aid learning, including planning, checking, monitoring, selecting, revising, and evaluating. These strategies are important components of models of self-regulated learning in the classroom, such as the model put forward by Pintrich (2004). Students' self-reported use of such strategies has been shown to correlate positively with academic outcomes (Pintrich & de Groot, 1990).

Researchers have long been interested in finding ways to help students acquire more complete and adaptive metacognitive skill in order to achieve better learning outcomes across academic topic areas and even across disciplines. In an influential volume entitled *How People Learn* (Bransford, Brown, & Cocking, 2000), one of three broad recommendations is to focus on improving students' metacognitive skills. A number of instructional programs focused on improving metacognition have been shown to be successful in actual classrooms. Some notable examples are programs that have focused on reciprocal teaching of reading skills (Palincsar & Brown, 1984; Rosenshine & Meister, 1994), on self-assessment of application of a scientific inquiry cycle (White & Frederiksen, 1998), and on strategies for self-regulation of mathematical reasoning (Mevarech & Kramarski, 2003). Classroom programs with a metacognitive focus have also been created for students with learning disabilities, although their effectiveness has not always been established in controlled experiments (Butler, 1998; Guthrie, Wigfield, & vonSecker, 2000).

In recent years, there has been increasing research on deploying metacognitive support

within computer-based learning environments, particularly within intelligent tutoring systems (Azevedo & Cromley, 2004; Arroyo et al., 2007; Biswas, Leelawong, Schwartz, Vye, & TAGV, 2005). The research presented in this chapter focuses on novel ways of supporting metacognition in real educational settings. This is done using computer-based tutoring of specific metacognitive skills, as contrasted to a comprehensive program of metacognitive instruction. In discussing various mechanisms to engage students in utilizing appropriate metacognitive strategies, we distinguish between static and adaptive metacognitive support (cf., Azevedo, Cromley, & Seibert, 2004). By static support for metacognition we mean prompts or scaffolds that are present throughout instructional activities and do not vary between students. Unlike Azevedo et al.'s (2004) notion of "fixed" support, our notion of static support may also involve feedback to the student as to whether they are successfully following the prompt. A classic example of static metacognitive support are prompts to students to "self-explain" (Chi, Bassok, Lewis, Reimann, & Glaser, 1989), which may occur in a text, a computer interface, or may be consistently delivered by a human tutor. If delivered by a computer or human tutor, these prompts may be followed by feedback to the student as to whether or not they successfully executed the prompt. Even when combined with feedback of this nature, we consider them to be static support, because the decision whether and when to prompt is not situation-specific, and not specific to any aspect of student metacognition. Adaptive support for metacognition on the other hand involves an active instructional agent, such as a teacher, human, or computer tutor, that implements instructional strategies that adapt to individual students' metacognitive behavior, for example, by fading prompts for metacognitive behaviors (i.e., gradually reducing the frequency of these prompts), or by providing feedback on metacognitive errors. This definition of adaptive metacognitive support is more specific than Azevedo et al.'s (2004) in that it specifically requires adaptation to student metacognition, not just to any aspect of student behavior. It thus requires that the system assesses aspects of student metacognition.

This chapter reviews four examples of our work on computer-based tutoring systems that provide either static or dynamic support for metacognition:

- 1 Static support for self-explanation (Alevén & Koedinger, 2002).
- 2 Adaptive support for error self-correction (Mathan & Koedinger, 2005).
- 3 Adaptive ways of reducing "gaming the system" (Baker et al., 2006).
- 4 Adaptive tutoring of help-seeking skills (Roll, Alevén, McLaren, & Koedinger, 2007a).

In evaluating the effect of these forms of metacognitive support, we have focused our efforts on assessing improvements in "robust learning." Robust learning produces a high level of conceptual understanding and/or procedural fluency so that students perform well not only on immediate post-tests with items highly similar to training but also on tests of transfer, long-term retention, and preparation for future learning (see learnlab.org/research/wiki/index.php/Robust_learning). As shown in Figure 20.1, we have defined four goals for robust learning of metacognitive skills, each of which builds on the previous one.

The first goal is for students to improve their metacognitive behavior within the learning environment while they receive metacognitive support. Ideally, this improvement will lead to better learning gains in the domain targeted by the supported environment, which is the second goal of metacognitive tutoring. The third goal is for students to internalize the metacognitive practices and thus to demonstrate better metacognitive behavior in subsequent instruction using a similar interface. Our fourth goal is that the metacognitive support will make students better at future domain-level learning (faster and more complete), based on the metacognitive practices they internalized.

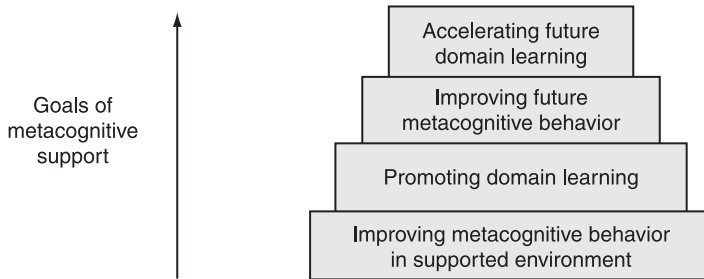


Figure 20.1 Goals of metacognitive tutoring. Toward the bottom are less ambitious and more proximal goals that support the more ambitious, ultimate goals above.

As our research has evolved, we have come to the view that an ideal experiment investigating the effect of metacognitive support will measure all four levels of learning. However, so far we have focused mainly on the first two goals; only in our most recent experiments on tutoring help-seeking have we addressed the third goal. The fourth goal remains for future work. In the first and second lines of research, focused on the metacognitive strategy of self-explanation (Chi et al., 1989) and error self-correction (Nathan, Kintsch, & Young, 1992), we assessed the effect of the metacognitive support on domain learning but not on students' metacognitive activities. In the third and fourth lines of research, focused on discouraging gaming the system and encouraging better independent help seeking, we not only assessed changes in domain learning, due to the metacognitive support, but also in the use of appropriate and inappropriate metacognitive strategies while the intervention is in place. In the experiment on tutoring help seeking, we also evaluated whether students' help-seeking behavior had improved following the intervention.

The reported research emphasizes controlled *in vivo* experimentation (Koedinger & Corbett, 2006; learnlab.org/research/wiki/index.php/In_vivo_experiment), that is, experiments or quasi-experiments that strive to maximize both internal validity and ecological validity. With respect to ecological validity, *in vivo* experiments use actual educational content and are run for realistic durations (hours rather than minutes of instruction) in the context of real courses and school settings. With respect to internal validity, *in vivo* experiments include a control condition and a treatment that involves a manipulation that targets a single causal principle (i.e., not full program variations as in typical randomized field trials) and whenever possible there is random assignment of students (or course sections) to condition. The “ecological” control condition used represents existing practice at the school setting.

Tutoring Self-Explanation

Our first line of research illustrating intelligent tutor support for metacognition focuses on the effect of supporting self-explanation. Of interest is a simple recurring metacognitive choice: After producing or being given a problem-solving step, students can either go on to the next step, or they can decide to self-explain the step they just completed, as a check of their understanding. Self-explanation has been studied extensively in the cognitive sciences, with one notable example being Chi et al. (1989), which reported an association between this metacognitive skill and learning. Static prompts for self-explanation have been shown in a number of lab studies to have a positive effect on student learning (Chi, de Leeuw, Chiu, & Lavancher, 1994; Renkl, Stark, Gruber, & Mandl, 1998; Siegler, 2002). Within the realm of intelligent tutoring systems, Conati and VanLehn (2000, 2001) implemented

dynamic prompts for self-explanations, which were presented based on the system’s assessment of the student’s domain knowledge and the student’s understanding of the example that they were explaining. They found that these prompts, combined with feedback on self-explanations, were superior to a self-explanation control condition that used static prompts, although this control condition differed in a number of other ways as well. These researchers did not investigate, however, whether self-explanation support improved learning over a non-self-explanation control condition, nor did they investigate whether students had become better self-explainers after the self-explanation support was removed. Our studies addressed the first limitation, whereas addressing the second remains for future work.

In two studies, we explored the effect of static support for self-explanation in an intelligent tutor that supports guided problem-solving practice. For each solved problem step, the tutor solicited simple self-explanations (namely, the name of the domain principle being applied, typically the name of a geometry theorem), and provided feedback on the correctness of these self-explanations. It did not, however, adapt to students’ level of domain knowledge, or metacognition, and thus this type of support does not qualify as adaptive support.

Our research into self-explanation started with a practical and widespread educational problem: students often develop shallow reasoning strategies as they learn in a new domain or attempt to learn a new cognitive skill. In the high-school geometry course that we were developing at the time, we observed that students often appear to draw on shallow strategies such as that illustrated in Figure 20.2. In this example, the student erroneously infers that two angles that look the same (namely, Angle 1 on the right and the un-numbered angle on the left) must have the same measures. We also found that the students were better at solving geometry problems than they were at justifying their reasoning steps by pointing out which theorems they were using (Alevan, Koedinger, Sinclair, & Snyder, 1998). This finding may be a manifestation of shallow reasoning, although other explanations are possible.

It is important to note that superficial strategies occur in many domains and with many forms of instruction. In physics problem solving, for example, there are many common misconceptions, such as confusing mass and weight, or the direction of velocity and

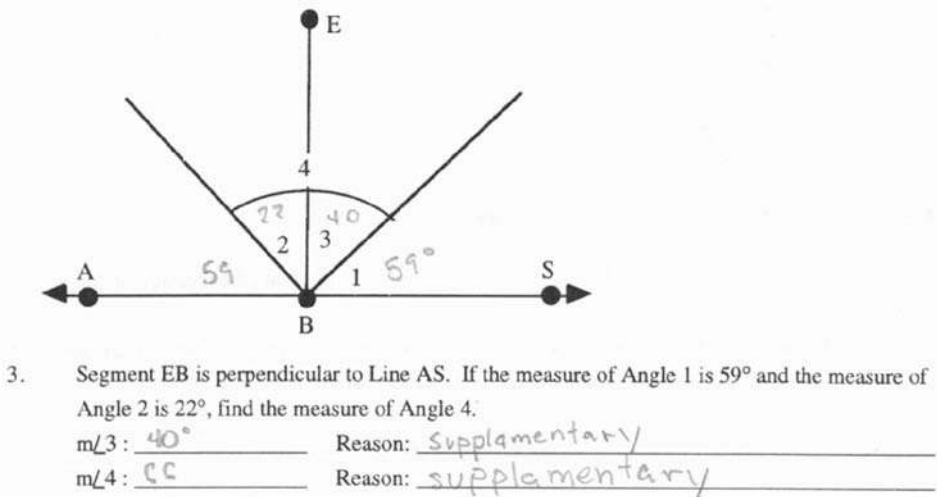


Figure 20.2 Shallow reasoning in geometry.

acceleration (Ploetzner & VanLehn, 1997). Novices often classify physics problems by superficial features that are not typically related to the solution method (Chi, Feltovich, & Glaser, 1981). In mathematics, students may learn procedures without ever grasping the underlying principles (Byrnes & Wasik, 1991; Fuson, 1990; Hiebert & Wearne, 1996). In the domain of writing, students often exhibit a “knowledge telling” strategy in which they simply list everything they know about a subject rather than constructing a well-organized argument (Scardamalia & Bereiter, 1985).

Studies to Evaluate Self-Explanation Support

We conducted two experiments to test the hypothesis that simple self-explanation support, requiring that students connect their answers to the underlying problem-solving principles, would help students to learn in a more robust manner. We investigated this hypothesis in the context of the Geometry Cognitive Tutor. We summarize the second of two experiments (Alevan & Koedinger, 2002).

The Geometry Cognitive Tutor is a form of intelligent tutoring software. It was developed in our lab and is being used in hundreds of schools across the United States. This software (as well as the Cognitive Tutors used in the studies reported in subsequent sections of this chapter) is based on the ACT-R theory of cognition and learning (Anderson & Lebière, 1998) as well as on a set of principles for the design of Cognitive Tutors (Koedinger & Corbett, 2006). The Cognitive Tutor software supports students as they acquire complex cognitive skill. It selects problems for students on an individual basis, and provides step-by-step guidance as students solve problems. The tutor’s guidance consists of contextual hints (given at the student’s request) about what to do next, feedback on correctness, and just-in-time messages that provide help with common errors. Cognitive Tutors have been shown to significantly increase high-school students’ math achievement compared to more typical classroom instruction (see the summary in Koedinger & Alevan, 2007).

The standard Geometry Cognitive Tutor served as the ecological control condition in the experiment, meaning that it represents standard practice in the field. We shall refer to it as the Problem Solving Condition. For the treatment condition (Explanation Condition) we created a version of this tutor that was enhanced to support self-explanation but was otherwise the same (see Figure 20.3). For each problem-solving step, the tutor prompted the student to provide a reason, namely, the name of the theorem being used. The students could type the name of the reason if they remembered it, or they could select it from the tutor’s on-line glossary of geometry knowledge, a resource with information about the approximately 25 different theorems being targeted in the given tutor unit. For each theorem, there was a complete statement expressing the theorem and a simple example of how the theorem could be used in solving problems. Students could double-click on glossary items in order to select them as the explanation for the current step. The tutor provided correctness feedback on the explanations, and if the student requested a hint, provided hints related to the correct explanation. The software required that a student get the explanations right before moving on to the next problem.

All students took a pre-test and post-test, with three kinds of test items. First, “Answer” items were designed to assess whether students could solve geometry problems of the same type that they encountered as they were working with the tutor. Second, “Reason” items were designed to assess whether students could explain their reasoning in the same manner they did when working with the tutor. The Answer and Reason items followed the same format as the corresponding steps in the tutor problems. (We note that the students in the Problem Solving Condition did not practice explaining their reasoning as they worked

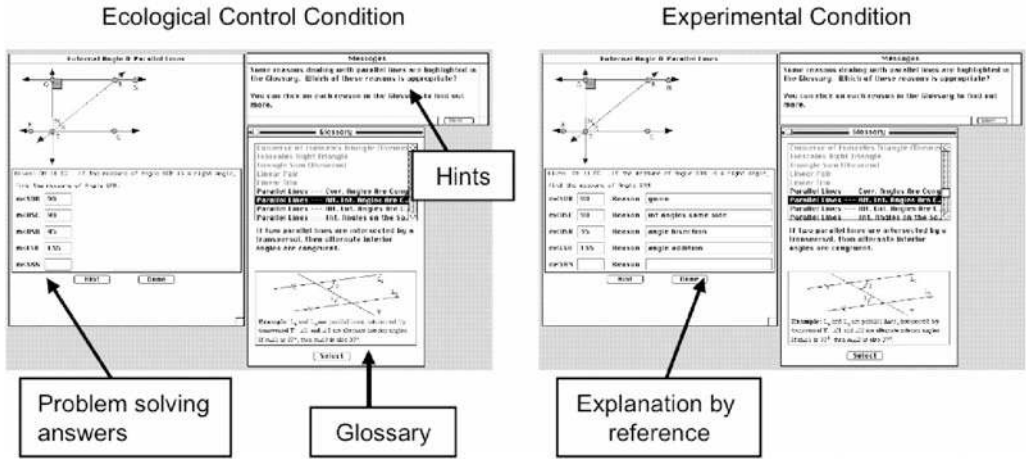


Figure 20.3 In the Experimental Condition, students explained their steps “by reference.” Otherwise, the tutor versions used in both conditions were the same.

with the tutor. For them, therefore, these post-test items were a form of transfer.) Third, “Not Enough Info” items tested whether students could recognize situations where there was insufficient information to infer the measure of a given quantity. These items were designed to expose shallow knowledge, because shallow strategies such as “if angles look the same, they are the same” will often lead to wrong answers on these items. These items provide a measure of transfer, because the students in neither condition had encountered these types of items during their work with the tutor.

We found that students who had explained their steps during their work with the tutor showed greater gains with respect to Not Enough Info items and with Reason items than the students who solved problems without explaining their steps (see Figure 20.4). The performance on the Answer items did not differ between the two conditions, despite the fact that the students in the Explanation Condition had encountered only half as many of them during training as their counterparts in the Problem Solving Condition. (The students in the Explanation Condition spent a significant portion of their time explaining steps. Because we controlled for time-on-task, these students completed fewer problems as they worked with the tutor.)

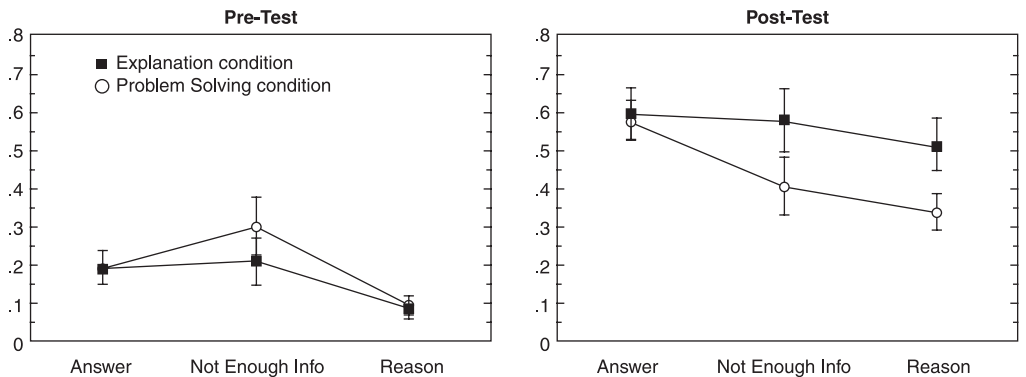


Figure 20.4 Pre- and post-test results for students in the second study, on each of the three types of test items. Multiple forms were used to counterbalance for test difficulty.

Thus, the students who had explained their steps learned in a more robust manner compared to students in an ecological control condition. They did better on two types of items that hinge on deeper understanding, namely Reason items and Not Enough Info items. The Reason items tap conceptual knowledge, and the Not Enough Info are transfer items designed to expose shallow knowledge. A further analysis suggested that students in the Explanation Condition acquired less shallow procedural knowledge and more flexible conceptual/declarative knowledge (Aleven & Koedinger, 2002).

Discussion of the Self-Explanation Support Studies

The experiment shows that supporting metacognition during instruction can improve robust learning of domain knowledge. We found that simple self-explanation prompting, with feedback on self-explanations, leads to robust learning. A limitation of the current study, addressed in some of the studies discussed below, is that we assessed domain (geometry) learning, not metacognitive improvement. We did not assess, for example, whether students were more likely to spontaneously self-explain at appropriate junctions during subsequent learning experiences. Nonetheless, the finding that even simple self-explanation support can be effective in intelligent tutoring software is of practical and theoretical interest, especially given that it shows improvement over a control condition that itself has been shown to improve upon typical classroom instruction (Koedinger & Aleven, 2007). Other researchers have since replicated the effectiveness of static menu-based support for self-explanation, combined with feedback on students' self-explanations (Atkinson, Renkl, & Merrill, 2003; Corbett, McLaughlin, Scarpinato, & Hadley, 2000; Corbett, Wagner, & Raspat, 2003).

Studies with interfaces that support free-form entry of self-explanations (i.e., students type in explanations) suggest that feedback on self-explanations may be an important factor. Aleven and colleagues created two versions of the Geometry Cognitive Tutor that requested free-form explanations from students, one that did not provide any feedback (Aleven & Koedinger, 2000b), and one that provided feedback through a natural language dialogue module (Aleven, Ogan, Popescu, Torrey, & Koedinger, 2004). In two studies, they compared these tutor versions against menu-based self-explanation (i.e., against the experimental condition in the experiment described above). They found that prompts for free-form explanation without feedback are ineffective. Students largely ignore them and provide very few good explanations (Aleven & Koedinger, 2000b). With feedback, however, free-form explanations help students learn to state better explanations, with no detrimental effect on problem-solving skill or transfer, even though self-explanations take time away from solving problems per se (Aleven et al., 2004). In addition, Corbett and colleagues, in a study with a Cognitive Tutor for Algebra II, found that free-form explanations with feedback led to slightly (but reliably) better transfer (Corbett, Wagner, Lesgold, Ulrich, & Stevens, 2006).

Researchers have begun to experiment with simple adaptive forms of self-explanation support, but it is too early to tell what works and what does not. The adaptive support in Conati and VanLehn's (2000) system was effective but the study conducted did not isolate the effect of the adaptive nature of the support. Hausmann and Chi (2002) reported that computer-based prompts are as effective as human prompts when students type self-explanations into a computer interface without feedback. The prompts were yoked across the conditions, so the study suggests that adaptive prompts may not be necessary, at least when students do not receive feedback on their self-explanations. A study by Weerasinghe and Mitrovic (2005) reported that self-explanation prompts selected based on students' domain knowledge improve domain-level learning. Conati, Muldner, and

Carenini (2006) point the way toward truly adaptive support. They developed a system that decides when to present prompts for self-explanation, and what type of self-explanation to prompt for, by trying to predict how much a given student will benefit from each type of self-explanation. In summary, a number of researchers are actively investigating whether adaptive self-explanation support can help students learn more robustly at the domain level and become better future learners.

Tutoring Error Self-Correction

Our second line of research illustrating intelligent tutor support for metacognition focuses on the effect of supporting students in error self-correction. This research program grew out of a reframing of the debate over feedback timing. Some researchers have demonstrated and argued for benefits of delayed feedback over immediate feedback in instruction and training (Schmidt & Bjork, 1992). An intuition behind this argument is that delayed feedback leaves room for the student to engage in self-detection and correction of their own errors. Other research (Corbett & Anderson, 2001) has shown benefits of immediate feedback, particularly in enhancing the efficiency of learning while maintaining learning outcomes. Mathan and Koedinger (2005) reframed this feedback debate by suggesting that perhaps the more relevant difference should be in the nature of the “model of desired performance” or the instructional objectives. One may want to produce error-free efficient expertise or, alternatively, one may want to produce flexible problem solvers (also called “intelligent novices”—Bruer, 1993) who may make some initial errors but who, unlike novices, eventually detect and correct these errors. In the context of an intelligent tutoring system, the immediate feedback given relative to an intelligent novice model of desired performance will actually be similar to delayed feedback given relative to an expert model of desired performance, as both will wait to give feedback on errors. However, there is one key difference. Namely, unlike delayed feedback, tutoring with the goal of producing intelligent novices involves providing assessment/monitoring, feedback, and hints at the metacognitive level that guide, as needed, the process of error detection and correction. That is, this kind of tutoring constitutes an *adaptive* form of metacognitive support, as defined above.

Error self-correction requires students to reflect on their performance, on the outcomes of that performance, how those outcomes are different from desired outcomes and, most importantly, on the rationale or reasoning for the initial performance attempt and how that reasoning might be modified in order to revise the performance and achieve the desired outcome. Other researchers have also focused on the metacognitive processes of monitoring one’s performance and progress, to check for errors or evaluate level of understanding (Flavell, 1979; Palincsar & Brown, 1984; Pintrich, 2004; Schoenfeld, 1983; White & Frederiksen, 1998). This project focused particularly on supporting reflective reasoning after an error has been identified. During such reflection learners think about their thinking—about *why* they decided to do something and about what might be wrong (and thus can be fixed) in that reasoning process. This study illustrates this process in the context of Excel spreadsheet programming and in particular the use of relative and absolute references (explained in Figure 20.5).

Notice in Figure 20.5a and 20.5b how when the formula is copied in cell B4, the formula produces \$6000 rather than the desired result of \$300. Excel performs “relative referencing” such that, for instance, if a referenced cell is above the formula cell in the original formula (as B2 is in the formula in B3), then it will remain above it in the copied cell (the reference will become B3 when the formula is copied into cell B4). In contrast, in Figure 20.5c, a correction to the formula is made employing “absolute

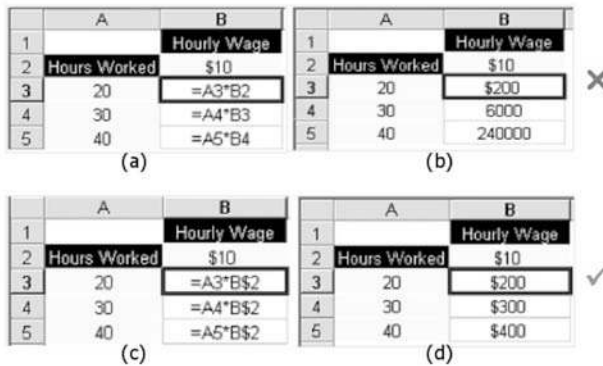


Figure 20.5 Two different attempts at the task of writing a formula to multiply a fixed hourly wage (in cell B2) by different hours worked (in cells A3, A4, and A5) to produce the amounts earned (cells B3, B4, and B5). Panel (a) shows the entry of an incorrect formula in cell B3 and what the resulting formulas look like after the original is copied into cells B4 and B5. When a formula in Excel is copied and then pasted in new cells, the cell references are updated “relative” to the original position. This produces the wrong result, as seen in panel (b). However, if a “\$” is placed in front of a column letter or row number (panel c), the reference is “absolute” and will stay the same when copied (when a reference to B\$2 is copied from B3 to B4, it does not change to B\$3, but remains B\$2). The result shown in panel (d) is now correct.

referencing” by adding “\$” in front of the 2 in B2. Thus, the correct result is achieved as shown in 5d. How might a learner go about reasoning reflectively about the error shown in 5b?

Experiments on Tutoring Self-Correction

Figure 20.6 illustrates the difference both in timing and content of the feedback and hints between the Expert (EX) tutor condition (see panels a and b) and the Intelligent Novice (IN) tutor condition (see panels c and d). The experiment performed, while not run in the context of a real course, had a number of features of *in vivo* experimentation including ecological validity factors: real content, appropriate students (temporary employment workers whose skills could be enhanced through participation), realistic duration (nearly 3 hours of instruction), and internal validity factors (single principle variation and random assignment). In addition to using a post-test immediately following instruction and using items isomorphic to the training, robust learning measures were also used including measures of long-term retention (one week later) and of conceptual and procedural transfer. The results demonstrated benefits for the Intelligent Novice tutor condition on all measures (Mathan & Koedinger, 2005). For instance, Intelligent Novice (IN) students performed significantly (and statistically reliably) better (at 74% correct) on the delayed transfer test than students in the Expert (EX) condition (at 60% correct). IN students were also significantly better (88% vs. 79%) on an immediate test involving Excel formula programming items that were isomorphic to those used in the instruction. Learning curve analysis of log data showed that the effect of the treatment made a difference quite early in the instruction (Martin, Koedinger, Mitrovic, & Mathan, 2005). This result suggests that error correction support was more relevant to students’ initial declarative understanding than later refinement through practice.

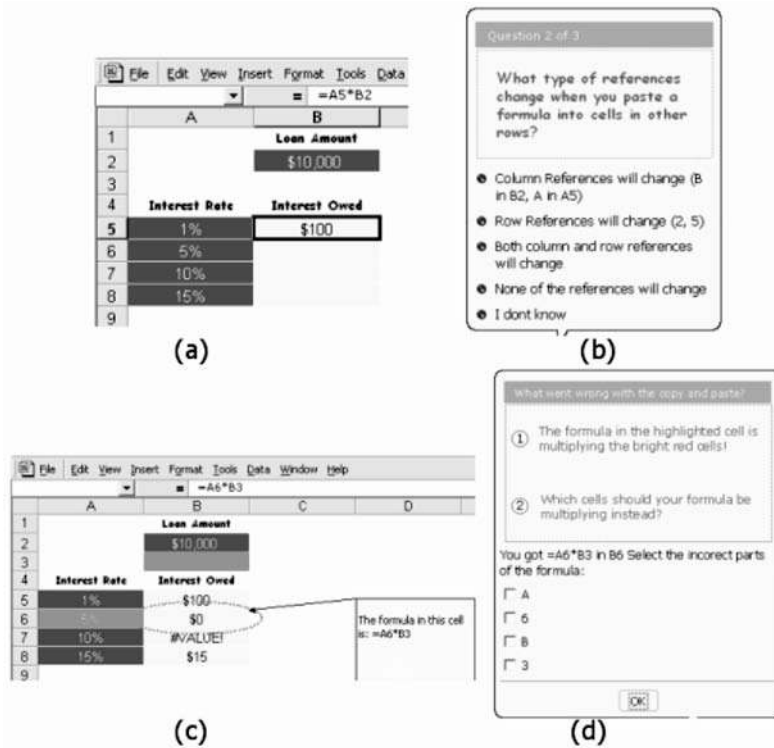


Figure 20.6 At the point of initial error (a), the Expert tutor provides hints as shown in (b). In the Intelligent Novice tutor, however, it is at the later point after copying and pasting the formula (c) that the Intelligent Novice hints are given (d). Note the different character of the hint messages (b) and (d). The Intelligent Novice hints (d) help students understand the reasoning behind a correct formula by analyzing what went wrong with incorrect formula entry.

Discussion of Tutoring Error Self-Correction

The early differences in favor of the IN condition (nominally delayed feedback) are inconsistent with predictions of Bjork’s (1994) “desirable difficulties” notion that suggests a trade-off whereby delayed feedback should hurt immediate performance (relative to immediate feedback), but enhance long-term learning. This inconsistency may be attributed to the fact that the Intelligent Novice intervention is not simply delaying feedback but providing direct assistance for error correction reasoning when needed. There may also be differences between Bjork’s prior results and ours because of the simpler, motor-skill-oriented domains in that work in contrast to the more complex and semantically-rich domain addressed here. The effectiveness of the intervention presented here appears to be in how it helps students reason to a better understanding of the required inferences and not in improving a general student ability to self-detect and fix errors. Metacognitive support from the tutor helps students to reason about their errors. By reasoning about correct procedures in contrast to incorrect ones, students appear to develop a better conceptual understanding of those procedures. We suspect that if students had simply been given delayed feedback, without the support for reasoning about error correction (as illustrated in Figure 20.5d), we would not see the positive outcomes we did. We do note that detecting and making sense of errors in this task domain appears relatively easy for students

(even if learning the procedures is not). Intelligent Novice error correction feedback may not work as well in other domains where error detection or error understanding is more difficult.

The Intelligent Novice feedback approach has interesting similarities to support for self-explanation in that it directs students toward the deep relevant features of the domain and away from the shallow irrelevant features (see learnlab.org/research/wiki/index.php/Features). Students are supported to think about what aspects of the formula may change in unwanted ways if an absolute reference (\$) is not used. Such a deeper encoding leads to better transfer than a surface or shallow encoding, for instance, thinking in terms of row and column movements and whether to put the \$ before the letter or the number. Helping students make the distinction between incorrect and correct performance seems to support deep encoding. It is similar to variations on self-explanation in which students are given an incorrect performance example (ideally a common or likely one) and asked to explain why it is wrong (Grosse & Renkl, 2007; Siegler, 2002).

Tutoring to Reduce Gaming the System

Tutoring metacognition not only requires guiding students to learn to choose appropriate learning and performance strategies, it also requires them to learn to not choose inappropriate learning and performance strategies. Consider a student using learning software, who does not know how to correctly solve the current problem step or task. This student might adopt the appropriate metacognitive strategy of seeking help from the software, the teacher, or another student, and then self-explaining that help. However, the student may instead choose a less appropriate strategy: “gaming the system.”

We define gaming the system as “attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material” (Baker et al., 2006). Gaming behaviors have been observed in a variety of types of learning environments, from intelligent tutors (Schofield, 1995), to educational games (Magnussen & Misfeldt, 2004) to online course discussion forums (Cheng & Vassileva, 2005). Gaming behaviors have been found to be associated with significantly poorer learning in intelligent tutors (Baker, Corbett, Koedinger, & Wagner, 2004). Additionally, there is some evidence that gaming the system on steps that the student does not know (termed “harmful gaming”) is associated with poorer learning outcomes than gaming the system on steps the student already knows (termed “non-harmful gaming”) (Baker, Corbett, Roll, & Koedinger, 2008a). Within intelligent tutoring systems, gaming the system generally consists of the following behaviors:

- 1 Quickly and repeatedly asking for help until the tutor gives the student the correct answer (Wood & Wood, 1999).
- 2 Inputting answers quickly and systematically. For instance, systematically guessing numbers in order (1,2,3,4 . . .) or clicking every checkbox within a set of multiple-choice answers, until the tutor identifies a correct answer and allows the student to advance.

In this section, we discuss a pedagogical agent, Scooter the Tutor, designed to reduce gaming and increase gaming students’ learning (Baker et al., 2006). Scooter the Tutor was built around a machine-learned detector of gaming behavior, which was shown to accurately determine how much each student games the system and on which problem steps the student games (Baker et al., 2008a). This detector was able to distinguish between the two types of gaming behavior mentioned earlier (harmful and non-harmful gaming), enabling

Scooter to focus solely on responding to harmful gaming, the behavior associated with poorer learning. This detector has also been shown to effectively transfer between different tutor lessons with little reduction in effectiveness (Baker et al., 2008a).

Scooter was designed to both reduce the incentive to game, and to help students learn the material that they were avoiding by gaming, while affecting non-gaming students as minimally as possible. Scooter was built using graphics from the Microsoft Office Assistant, but with modifications to enable Scooter to display a wider range of emotion. During the student's usage of the tutoring system, Scooter responds to gaming behavior in two ways: via emotional expressions and supplementary exercises.

When the student is not gaming, Scooter looks happy and occasionally gives the student positive messages (see the top left of Figure 20.7). Scooter's behavior changes when the student is detected to be gaming harmfully. If the detector assesses that the student has been gaming harmfully, but the student has not yet obtained the answer, Scooter displays increasing levels of displeasure (culminating in the expression shown on the bottom left of Figure 20.7), to signal to the student that he or she should now stop gaming, and try to get the answer in a more appropriate fashion.

If the student obtains a correct answer through gaming, Scooter gives them a set of supplementary exercises designed to give them another chance to cover the material bypassed by gaming this step, as shown in Figure 20.7. Within supplementary exercises, the student is asked to answer a question. This question may require understanding one of the concepts required to answer the step the student gamed through, or may require understanding the role the gamed-through step plays in the overall problem-solving process. If the student tries to game a supplementary exercise, Scooter displays anger.

Study on Effects of Tutoring Students not to Game the System

We studied Scooter's effectiveness in an *in vivo* experiment in the context of a year-long Cognitive Tutor curriculum for middle-school mathematics (Koedinger & Corbett, 2006), within five classes at two schools in the Pittsburgh suburbs. The study was conducted in

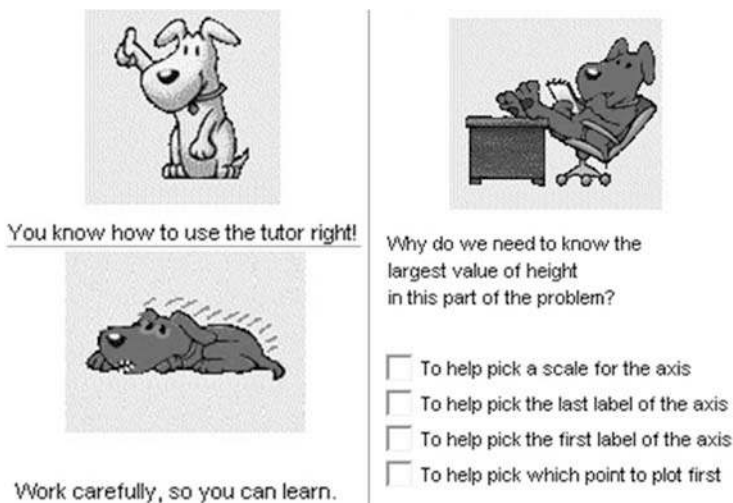


Figure 20.7 Scooter the Tutor—looking happy when the student has not been gaming harmfully (top left), giving a supplementary exercise to a gaming student (right), and looking angry when the student is believed to have been gaming heavily, or attempted to game Scooter during a supplementary exercise (bottom left).

the spring semester, after students had already used the Cognitive Tutor for several months. Each student used a tutor lesson on scatterplots. The control condition and experimental condition occurred during different weeks—students not using the scatterplot tutor used a different tutor lesson on another subject. Fifty-one students participated in the experimental condition for the scatterplot lesson (12 were absent for either the pre-test or post-test, and thus their data will not be included in analyses relevant to learning gains); 51 students participated in the control condition for the scatterplot lesson (17 were absent for either the pre-test or post-test).

Before using the tutor, all students first viewed instruction on domain concepts, delivered via a PowerPoint presentation with voiceover and simple animations. In the experimental condition, a brief description of Scooter, and how he would respond to gaming, was incorporated into the instruction. Then students completed a pre-test, used the tutor lesson for 80 minutes across multiple class periods, and completed a post-test. Test items were counter-balanced across the pre-test and post-test. Log files were used to distill measures of Scooter's interactions with each student, including the frequency with which Scooter got angry, and the frequency with which Scooter gave a student supplementary exercises.

Observational data was collected to determine each student's frequency of gaming, using quantitative field observations (systematic observations of student behavior by field observers, cf. Baker et al., 2004), in order to analyze Scooter's effects on gaming frequency. Another potential measure of how much each student gamed, the gaming detector, was not used because of risk of bias in using the same metric both to drive interventions and as a measure of the intervention's effectiveness.

Results of Intervention on Metacognitive Behavior and Learning

The Scooter intervention was associated with a sizeable, though only marginally significant, reduction in the frequency of observed gaming; 33% of students were seen gaming in the control condition (using quantitative field observations), while 18% of students were seen gaming in the experimental condition. However, although fewer students gamed, those students who did game did not appear to game less (14% in the experimental condition, 17% in the control condition).

In terms of domain learning, there was not an overall between-condition effect. However, only a minority of students received a substantial number of supplemental exercise interventions from Scooter (because only a minority of students gamed the system, as in previous studies of gaming). There is some evidence that the intervention may have had an effect on these specific students. In particular, the supplemental exercises appeared to be associated with significantly better domain learning. The third of students (out of the overall sample) that received the most supplementary exercises had significantly better learning than the other two-thirds of the students, as shown in Figure 20.8. Students who received the most supplementary exercises started out behind the rest of the class, but caught up by the post-test (see Figure 20.9 Left). By contrast, in both the control condition (see Figure 20.9 Right) and in prior studies with the same tutor, frequent harmful gaming is associated with starting out lower than the rest of the class, and falling further behind by the post-test, rather than catching up.

The emotional expressions, on the other hand, were not associated with better or worse learning. Students who received more expressions of anger did not have a larger average learning gain than other students.

There was no evidence that students reduced their degree of gaming after receiving either type of intervention (according to the quantitative field observations). Hence, the

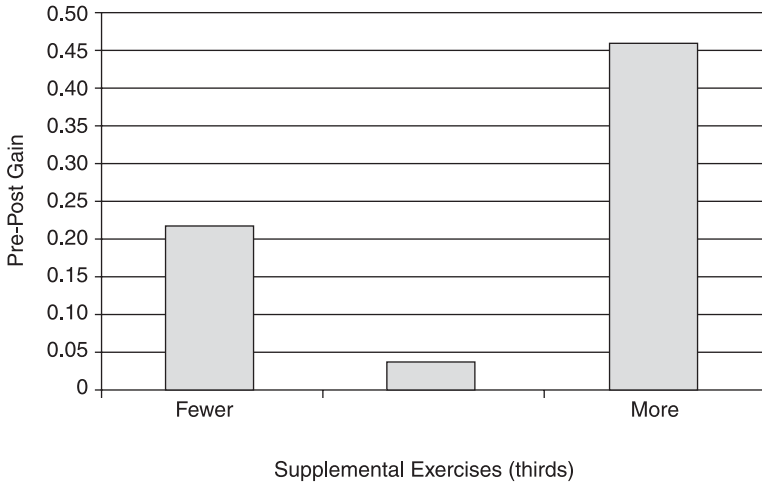


Figure 20.8 The learning gains associated with receiving different levels of supplemental exercises from Scooter.

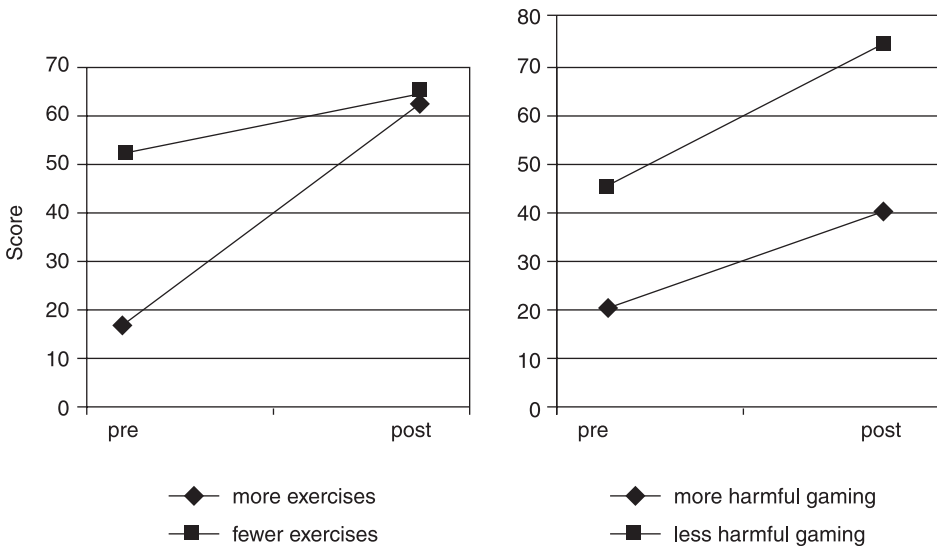


Figure 20.9 Left: The pre-test and post-test scores associated with receiving different levels of supplemental exercises from Scooter (top third versus other two-thirds). Right: The pre-test and post-test scores associated with different levels of harmful gaming in the control condition (top half of harmful gaming versus other students).

observed reduction in gaming may have been from Scooter’s simple presence. Students who chose to game knowing that Scooter was there did not appear to reduce their gaming.

Given the connection between receiving the supplementary exercises and learning, it is surprising that there was not an overall learning effect for Scooter. One potential explanation is that students who ceased gaming chose other ineffective learning strategies instead. The lack of connection between reduced gaming and improved learning may indicate that gaming is not directly causally related with learning. The apparent benefits of supplementary exercises may be from the “variable encoding” (Paas & Van Merriënboer, 1994) that

students experienced in the different kinds of presentation of the target knowledge in the original tutor and the supplementary exercises.

Recently, there have been two other systems designed to reduce gaming and improve gaming students' learning. Walonoski and Heffernan (2006) displayed sophisticated visualizations about gaming behavior to both students and teachers, on the student's screen. Arroyo et al.'s (2007) ProgressTips gave detailed metacognitive messages to gaming students about appropriate metacognitive behavior between problems. Walonoski and Heffernan's gaming visualizations, like Scooter, reduced gaming. Gaming was even reduced in future units where gaming visualizations were not given, showing some durability of the result. However, Walonoski and Heffernan did not measure domain learning. Arroyo et al.'s ProgressTips did not reduce gaming—instead, they caused students to shift gaming behavior (personal communication, Ivon Arroyo), as in Murray and VanLehn (2005). Domain learning, however, was improved by ProgressTips. ProgressTips also improved students' attitudes towards the tutor and learning domain, a positive effect not obtained with Scooter, who was generally disliked by students who received his interventions. One possible explanation is that ProgressTips disrupted students' learning experiences less than Scooter, who interrupted student behavior as soon as a student completed a step through gaming.

In general, the results of studies on these three systems to respond to gaming suggest that interventions given for differences in students' metacognitive behaviors can reduce the incidence of inappropriate behaviors and improve learning, whether they are given at the time of behavior or shortly afterwards.

Tutoring Help Seeking

The ability to seek help adaptively is a key metacognitive skill that figures prominently in theories of self-regulated learning (Newman, 1994), specifically as an important resource management strategy (Pintrich, 1999). Research on help seeking in social settings (e.g., classrooms) shows that adaptive help-seeking behavior can be an effective route to independent mastery of skills (Karabenick & Newman, 2006). Intelligent tutoring systems are an interesting context to investigate help seeking because they typically have sophisticated on-demand help facilities and they offer an opportunity for very detailed analysis of the usage patterns of these facilities. As part of the step-by-step guidance that these systems offer, students can typically request multiple levels of tailored help on what to do next at any point during their problem-solving activities. This form of help is thought to be beneficial for learning (e.g., Wood & Wood, 1999). As mentioned, the particular system that we worked with, the Geometry Cognitive Tutor, also provides an on-line glossary of geometry knowledge that lets students browse descriptions and examples of the geometry theorems that they are learning.

Upon closer scrutiny, however, the assumption that students' learning benefits from using these facilities turns out to be problematic. We found that students often do not use help facilities effectively (Aleven & Koedinger, 2000a; Aleven, Stahl, Schworm, Fischer, & Wallace, 2003). This finding mirrors findings by researchers who study help seeking in social settings (Arbreton, 1998; Newman, 2002). With surprising frequency, students abuse the tutor's help, focusing on help levels that essentially give away the next step, all but ignoring the help levels that provide explanations of why the answer is the way it is.

This behavior appears to generally be maladaptive, metacognitively, and has been shown to be associated with poorer learning (Aleven, McLaren, Roll, & Koedinger, 2006; Baker et al., 2004, 2008a). However, it appears that for a small number of students the rapid clicking through hints represents a way of efficiently turning a problem step into an

example step. The student may then self-explain the step, in the best case reconstructing the tutor's explanations. Providing support for this explanation, a recent data mining study examining tutor logs found that spending large amounts of time on a bottom-out hint is positively correlated with learning (Shih, Koedinger, & Scheines, 2008).

Another form of maladaptive help-seeking behavior seen in our research is when students resist using help even when they clearly seem to need it, such as after multiple errors on a step. For example, Alevan and Koedinger (2000a) found that after several consecutive errors on a step with no hints, students were more likely to try again rather than ask for a hint. In addition, the students used the tutor's glossary very rarely. On the other hand, when requesting a hint, students asked to see the bottom-out hint on 82% of their hint episodes. To summarize, students either ask for "instant" help or no help at all, but tend to avoid more complex help-seeking episodes.

Studies to Evaluate Tutoring of Help Seeking

Given the high frequency of maladaptive help seeking, we embarked on research to test the hypothesis that a Cognitive Tutor agent that provides guidance with respect to students' help-seeking behavior can help students to both learn better at the domain level and become better help seekers. In other words, this fourth project focuses on improving one aspect of students' metacognitive abilities, their ways of seeking help, and thereby their current and future learning. In this sense, this project is more ambitious than the other three projects described above: the goal is not just to "channel" students into a particular metacognitive behavior with the tutor, but to help students internalize a metacognitive behavior that transfers to future tutor use and even to other learning environments. Furthermore, the help-seeking project avoids using any domain-specific assumptions or concepts, and thus maintains a metacognitive character, making it possible to apply the underlying model of help seeking to different domains without much adaptation. That is, this project offers externally regulating tutoring with the goal of helping students internalize the productive strategies and better self-regulating their learning in the supported environment and beyond. In doing so, it targets all four goals as described in Figure 20.1.

As a first step, we developed a model that aims to capture both effective and ineffective help-seeking behavior (Alevan et al., 2006). In contrast to the previous project (Scooter), in which a model was built using machine learning, the model was built by hand, and implemented as a set of production rules. Lacking the guidance of a detailed normative theory of help seeking, we made extensive use of student-tutor log data and theoretical cognitive task analysis to design the model. For example, the initial version of our model prescribed that students should always take their time before attempting a step. When mining log data of student-tutor interactions, we noticed that fast actions by students on steps on which they are skilled correlate with large learning gains, a finding clearly at odds with our initial model. We therefore updated the model to allow fast attempts (including incorrect ones) on steps where a student is skilled, subject to limitations such as number of overall errors on this step. The model (as summarized in flow-chart format in Figure 20.10), stipulates that students should work deliberately, spending adequate time reading problem statements and hints, and that they should use help in one of three cases: when steps are not familiar to them; when they do not have a clear sense of what to do; and when they have made an error (as indicated by the tutor's domain-level feedback) that they do not know how to fix. The choice of what source of help to use, the tutor's on-demand hints or the glossary, should be driven by the student's self-assessed level of relevant knowledge. The less familiar a step is, the more contextualized the help requested should be. (We view the glossary as decontextual help and the hints as highly contextual help.)

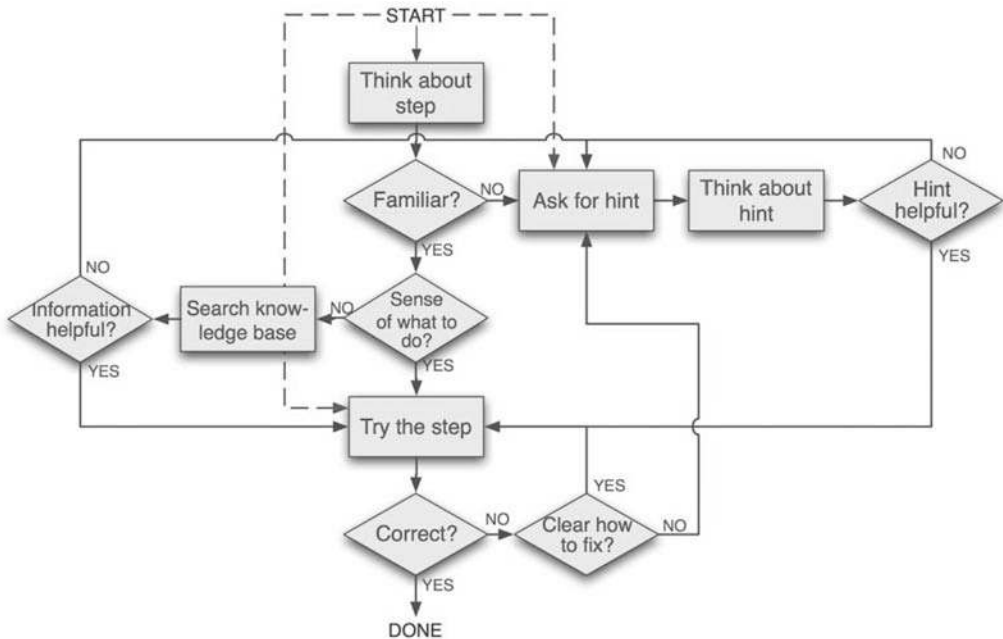


Figure 20.10 Schematic view of the help-seeking model.

Essentially, this model is a detailed normative theory of help seeking with an intelligent tutoring system. It specifies in a detailed, moment-by-moment manner what a student should do in any given situation. Notably, the model does not limit the student to a single learning trajectory. Rather, it allows for a wide variety of reasonable actions at each point, while excluding actions that were clearly shown to be ineffective in our modeling and data mining activities.

Data from our studies suggest that the help-seeking behavior captured by the model is associated with higher learning gains. In several analyses we have conducted, students who demonstrated poorer help-seeking behavior (i.e., students whose help-seeking behavior conformed to the model) were found to have poorer pre–post learning gains, though the result was not entirely stable across data sets (Roll, Baker, Alevan, McLaren, & Koedinger, 2005).

There is evidence that the model captures help-seeking behavior that is independent of the specific content or group of students. The model correlates well across different cohorts of students and tutors within the Cognitive Tutor family (Roll et al., 2005). We found that students tend to make the same types of help-seeking errors between Cognitive Tutors—the correlation between students’ frequency of different types of help-seeking errors in an angles unit (in geometry) and a scatterplot unit (in data analysis) was 0.89.

Next, we created the “Help Tutor” agent, a Cognitive Tutor at the metacognitive level. Driven by the help-seeking model, it provides context-sensitive feedback on students’ help-seeking behavior, as they work with a Cognitive Tutor. Student behavior that, in a given context, matches the normative predictions of the metacognitive model specific to that context, is deemed metacognitively correct. Student behavior that matches any of the many metacognitive “bug rules” in the model (which capture what is believed to be inappropriate help-seeking behavior) were deemed metacognitively incorrect. The Help Tutor displays a feedback message in response to such errors, as shown in Figure 20.11,

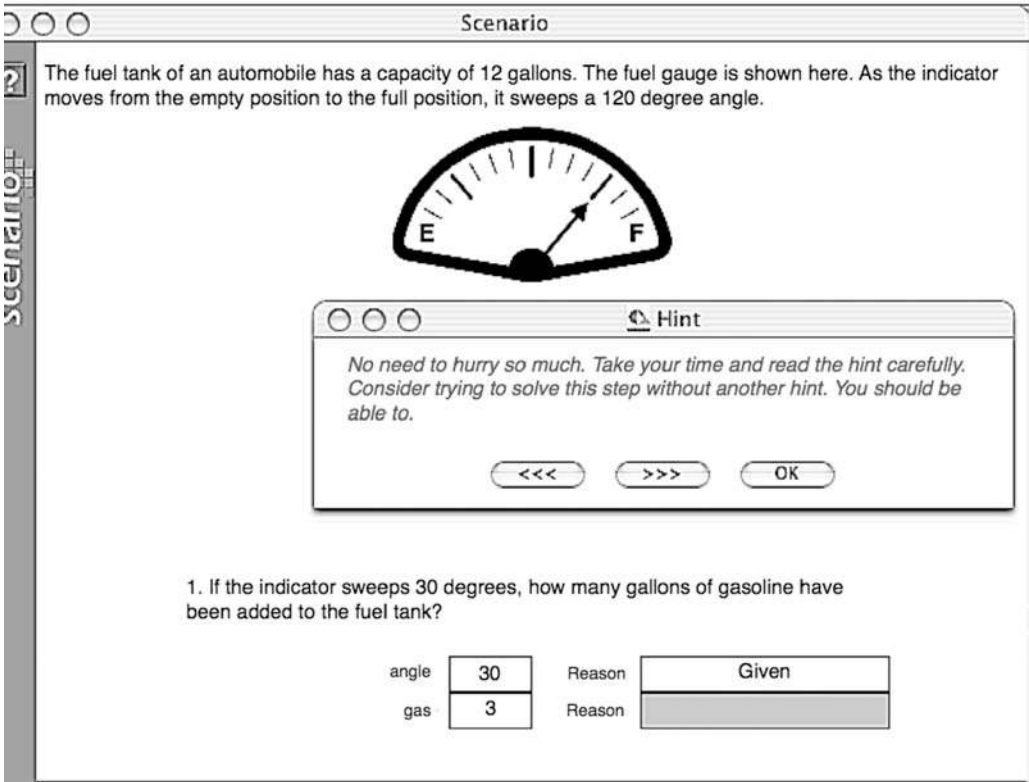


Figure 20.11 Example of a help-seeking error message.

“Perhaps you should ask for a hint, as this step might be a bit difficult for you,” or “It may not seem like a big deal, but hurrying through these steps may lead to later errors. Try to slow down.” Perhaps somewhat unfortunately, in retrospect, while the Help Tutor pointed out help-seeking behavior that was not likely to be productive, it never praised students for being good help seekers.

We integrated this metacognitive tutor agent into the Geometry Cognitive Tutor, so that students received guidance both with respect to geometry and with respect to their help-seeking behavior. The cognitive and metacognitive agents were not always in agreement; for example, an answer may constitute an error at the domain level, but attempting the step was still a metacognitively-appropriate action; likewise, a student may obtain a correct answer by guessing, metacognitively-inappropriate behavior. A prioritizing algorithm was implemented to choose the more informative feedback in cases where both cognitive and metacognitive feedback was appropriate.

We conducted two classroom experiments to evaluate the effects of the kind of metacognitive feedback generated by the Help Tutor (Roll et al., 2007a). The first study compared the traditional Geometry Cognitive Tutor to a version that included the Help Tutor, integrated with the Cognitive Tutor as described above. In the second study the Help Tutor was the main component of a broader metacognitive instructional package that also included declarative classroom instruction on help-seeking principles and short self-assessment activities supported by automated tutors. The classroom instruction included short video segments illustrating how to use the tutor’s help facilities effectively. In the self-assessment activities, the students were asked to rate their ability to apply a new

geometry theorem prior to solving their first problem involving that theorem, and then were asked to reflect on the correctness of their prediction (Roll et al., 2007a). In both experiments, the students in the control condition worked with the standard Geometry Cognitive Tutor. In both experiments, we tested the hypotheses that the respective metacognitive instruction (the Help Tutor in the first experiment, and the metacognitive instructional package in the second experiment) would lead to more desirable help-seeking behavior, both during practice with the tutor (goal 1 in Figure 20.1) and in a paper-and-pencil transfer environment after tutor usage (goal 3 in Figure 20.1), and that it would lead to better domain-level learning (goal 2 in Figure 20.1).

The studies showed mixed results. The evidence seemed to confirm the first hypothesis, that help-seeking behavior in the tutor would be improved: there was evidence that students sought help more appropriately under the Help Tutor's tutelage, as measured by the percentage of actions that conformed to the help-seeking model. However, the improvement was seen with respect to only a subset of help-seeking action types. There was no difference between the conditions in terms of help-seeking choices made on the first action on a new problem step before any Help Tutor feedback was seen. This finding may suggest that the improvement in students' metacognitive behavior was mainly the result of complying with the Help Tutor messages rather than of students' internalizing the metacognitive support and making proactive choices.

With respect to the second hypothesis, we found no lasting effect on students' help-seeking procedural knowledge, as measured by post-test scores on items with embedded hints, compared to performance on items with no embedded hints. We did, however, find that students who used the Help Tutor had a better declarative understanding of help-seeking principles, as measured by the quality of their answers to hypothetical help-seeking scenarios.

The evidence did not support the third hypothesis, that the metacognitive instruction would lead to better domain-level learning. Although students in both conditions improved significantly from pre- to post-test in both studies with respect to their geometry knowledge and skill, we did not find any differences between the conditions.

Discussion of Studies of Tutoring Help Seeking

In sum, only one of the three hypotheses was confirmed, namely, that the Help Tutor would lead to more productive help-seeking behavior. However, these studies have several limitations. For one, we did not measure help-seeking behavior in the subsequent unit within the same tutoring environment. That is, we did not address the fourth of the goals of metacognitive tutoring, depicted in Figure 20.1. While the improved help-seeking behavior did not transfer to the paper-and-pencil environment, it may have transferred to the subsequent unit within the Geometry Cognitive Tutor, which uses similar interface elements, problem types, learning goals, and requires similar learning strategies (as in Walonoski & Heffernan, 2006).

Interpretation of the results should take into account the context of the metacognitive instruction. Our studies focused on a specific type of learning environment, one that supports step-by-step problem solving with immediate feedback and hints available on demand. Such environments may encourage certain types of metacognitive behaviors and discourage others. For example, in the Cognitive Tutors we used, domain-level knowledge tracing (i.e., the Bayesian method by which the tutor tracks the student's detailed knowledge growth over time; Corbett & Anderson, 1995) uses only information from students' first action on each problem step. Therefore, students may be tempted to always enter the step before requesting help, even if they are essentially guessing, because they will receive

full credit for a correct answer, and if their answer is wrong they will not be penalized more than if they had requested a hint. These factors may promote an interaction style on the part of students that is specific to the particular learning software. It may be that in other environments, in which the advantages of productive help-seeking strategies are more apparent to students, help-seeking instruction will be more readily internalized by students.

Students' attitudes towards the Help Tutor may also help explain their behavior. Based on several interviews we conducted, and based on informal feedback, it appears that students did not like the system's commenting on their help-seeking behavior, even though they often agreed with the feedback. (A similar pattern was also observed with Scooter the Tutor, described above.) Even though students believed that the Help Tutor's suggestions were probably correct, they did not see its advice as valuable, and thus they may have merely complied with it, rather than internalizing the advice. It could be said that the students applied their existing help-seeking strategies to the Help Tutor itself. They were used to ignoring intermediate hints, and thus tended to ignore the Help Tutor. (This explanation is consistent with the lack of metacognitive improvement on first attempts before any Help Tutor feedback.)

The results of our experiment strongly suggest that many students, in spite of being aware of appropriate help-seeking strategies, choose not to apply those strategies. The students appeared to know the difference between ideal and faulty help-seeking behavior: they reported agreeing with the Help Tutor comments and, compared with the control group students, demonstrated better conceptual understanding of help seeking following usage of the Help Tutor and receiving the declarative help-seeking instruction. But their actual help-seeking behavior, on the post-test, was no better than the control condition students' help-seeking behavior. It did not reflect their superior declarative knowledge of help seeking. Apparently, students did not *want* to apply this knowledge. Hence, in order to create a learning environment that effectively promotes help seeking, we need to better understand the motivational and affective factors that shape students' help-seeking behavior and their desire to be (or become) effective help seekers and effective learners.

Relations Between Gaming and Affect, Motivation, and Attitudes

Why do students choose to engage in inappropriate learning behaviors or strategies? In particular, why do students choose to game the system, which is clearly an ineffective learning strategy? Answering this question may help in the development of future interventions that address gaming, in a more effective fashion than work up to this point has.

This broad question led us to investigate the relationship between gaming the system and several potential factors that could influence the choice to game, including affect, motivation, attitudes, and differences between tutor lessons. In this section, we briefly summarize six studies that we have conducted, in collaboration with our colleagues. These studies are reported in full detail elsewhere (Baker, 2007; Baker, Walonoski, Heffernan, Roll, Corbett, & Koedinger, 2008b; Rodrigo et al., 2007, 2008). In these studies, we correlated the frequency of gaming the system (measured either by behavioral observation or the gaming detector) and data from questionnaires and affect observations, with junior high school and high school students in the USA and Philippines. Five of six studies were conducted using intelligent tutoring systems—the sixth study involved an educational game.

One of the most common hypotheses for why students game (e.g., Martinez-Mirón, du Boulay, & Luckin, 2004; Baker et al., 2004) is that students game because they have

performance goals rather than learning goals. Two studies investigated this hypothesis—both found no such relationship, a surprising result especially given the finding by Pintrich (1999) that performance goals (termed extrinsic goals in this work) are associated with self-reports of not using self-regulated learning strategies. Items that indicated that the student had relatively low desire to persist in educational tasks were weakly associated with gaming, with r^2 under 0.05.

Another popular hypothesis among teachers (Baker et al., 2008b) is that gaming the system is associated with anxiety—however, two studies found no relationship between gaming and anxiety.

Negative attitudes towards computers, the learning software, and mathematics were each found to be correlated with gaming in one to three of the studies. These effects were statistically significant, but were fairly weak, with r^2 under 0.05.

The affective states of boredom and confusion were associated with the future choice to game. In two studies, a student's affect and behavior were repeatedly observed. In both studies, there was evidence that a bored student was more likely (about twice as likely) to begin gaming in the next three minutes, particularly among students who were frequently bored. There was a trend towards more gaming after confusion within one study, and a trend in the opposite direction in the other study. Gaming the system's future probability was not significantly increased by other affective states, including frustration, delight, surprise, and the flow state.

Hence, there were some fairly solid relationships between momentary affect and gaming, but only weak relationships between relatively stable attitudinal/motivational constructs and gaming. This finding led us to study, using data mining, how predictive these semi-stable student attributes are. The detector of harmful gaming was applied to each student's behavior in each tutor unit in an entire year of middle school-level tutor data, and the amount of variance in the student terms predicted as a class was used as a proxy, and an upper bound, for the total amount of variance that could be predicted by the sum of all student attributes that remain stable over a year. The result indicated that differences between lessons were much better predictors of how much a student would game ($r^2 = 0.55$) than differences between students ($r^2 = 0.16$).

Overall then, these results suggest that several factors contribute to a student's decision to game the system. Semi-stable student characteristics, in particular attitudes towards the domain, play a minor role in the choice to game the system. A larger role appears to be played by affect, with a student's momentary experience of boredom or confusion leading students in many cases to game the system shortly afterwards. This suggests that affective learning companions (Kort, Reilly, & Picard, 2001) that respond effectively to these affective states may in some cases be able to prevent gaming the system from occurring—it remains for future work to determine which affective responses might be effective in this regard. Another important factor is differences between tutor lessons—studying in greater depth how the differences between lessons increase or decrease gaming, and whether this effect is mediated by affect, may also help us to develop tutor lessons that students do not choose to game.

General Discussion and Conclusion

Within this chapter, we have discussed our research group's work to build tutorial support for metacognition, giving examples of interventions in the following four areas:

- 1 Self-explanation
- 2 Error self-correction

- 3 Reduction of gaming the system
- 4 Help-seeking skills

While we explored these metacognitive abilities within the context of intelligent tutoring systems, the results of the experiments may be relevant to other forms of instructional support for metacognition. The first two systems have achieved reliable domain-level effects on robust learning measures. The last two projects, on the other hand, had more modest or unstable effects on learning. In this discussion, we raise several hypotheses to explain the differences found between the outcomes of the given support and highlight key achievements and challenges.

Analysis of the Interventions

The four interventions differ along a number of dimensions, most notably in terms of how adaptive their metacognitive support is. As mentioned earlier, static support does not change depending on the students' behavior or knowledge; instead, it occurs during certain fixed stages in the learning task, regardless of student metacognition. In contrast, adaptive metacognitive support assesses aspects of students' metacognitive behavior and adapts the software's behavior or response, based on its assessment of the student's metacognitive behavior. Systems that offer adaptive support for metacognition typically allow students the freedom to commit metacognitive errors, and respond to these errors when they recognize them. Such systems may of course also respond when students engage in metacognitive behaviors deemed to be desirable or effective.

Under this definition, the approach to tutoring self-explanations discussed above is a form of static metacognitive support. In this intervention, the key metacognitive decision, namely, when/whether to self-explain one's problem-solving steps, is performed by the system not by the student, and consequently the system cannot assess students' metacognitive decisions. The system *does* assess the correctness of students' self-explanations, but once the decision to self-explain is made, the self-explanations themselves may well be viewed as domain-level behavior. Nevertheless, this tutor should still be viewed as supporting metacognition. Because it scaffolds students in engaging in positive metacognitive behavior, there is an opportunity for students to internalize this behavior and, perhaps more importantly, benefit from it at the domain level.

By contrast, the other three systems are adaptive to students' metacognitive behaviors. All three of them assess aspects of students' metacognitive behavior and respond according to their respective metacognitive assessment of the student, although they differ substantially in the range and type of metacognitive behaviors to which they respond. The Error Correction tutor recognizes a single desirable metacognitive behavior or lack thereof, namely, whether students correct their own (domain-level) errors as soon as the negative consequences of those errors are apparent. Further, this error correction recognition was only implemented for a limited (though carefully chosen) subset of the domain skills, namely formula writing errors. Scooter distinguishes a single broad undesirable metacognitive category, namely, harmful gaming. This category encompasses a range of more specific behaviors, such as guessing answers or help abuse, but Scooter does not "diagnose" them separately. It assesses metacognition at a more aggregate level than the other two systems. The Help Tutor, on the other hand, assesses students' help-seeking behavior (and also how deliberately students work with the Cognitive Tutor) in a highly fine-grained manner. It recognizes several different specific metacognitive errors, unlike the other three systems.

The systems differ further with respect to the content of the feedback they provide on

students' metacognitive behaviors. In any given system, the content of the feedback may relate primarily to the domain level, it may be primarily metacognitive, or it may involve both. As a further distinction, the feedback may comprise correctness information only, meaning that it indicates only whether the student engages in metacognitive behaviors deemed to be productive or not (without explaining why), or it may comprise more elaborate messages, for example, messages relating the student's behavior to specific principles of metacognition (e.g., help-seeking principles) or problem-solving principles at the domain level.

The intervention designed to tutor self-explanations provides feedback on the correctness of students' self-explanations. As mentioned, we consider this to be domain-level feedback. The intervention designed to tutor error correction reacts to metacognitive errors (namely, students not fixing a domain-level error when the consequence of that error is apparent) mainly in a domain-related way. It helps the student to generate correct behavior through a domain-specific discussion of what went wrong. It does not give students metacognitive level instruction, for instance, by suggesting a general strategy for checking for errors like attempting the problem again with a different strategy. Likewise, Scooter's main reaction to gaming the system is to assign remedial problems that address domain-level topics that the student is struggling with, according to its student model. These remedial problems constitute a form of domain-related feedback. Scooter also provides correctness feedback at the metacognitive level, both through brief messages and the animated animal cartoon character's emotional expressions. Of the four systems, the Help Tutor is the only one whose feedback is entirely metacognitive in content. It provides specific metacognitive error feedback messages, relating students' behavior to desired help-seeking principles.

Thus, the four interventions exhibit significant variability in their design. They provide case studies exploring different points in a design space for metacognitive support, defined by the dimensions discussed above, such as (a) whether the support for metacognition is static or adaptive, (b) the range and type of metacognitive behaviors that they respond to, (c) the level of detail at which they analyze students' metacognitive behaviors, (d) whether the feedback they provide in response to students' metacognitive behaviors is domain-related or metacognitive in nature, and (e) whether it is correctness feedback only or whether it is elaborate and explanatory. We expect that further dimensions in this design space for metacognitive support will be identified as researchers continue to develop and evaluate methods for automated tutoring of metacognition.

Although we cannot be said to have systematically explored the space defined by these dimensions, and it is therefore not possible to draw definitive conclusions about the effectiveness of any individual dimension or design feature, it is nonetheless interesting to ask what hypotheses we can "extract" about the kinds of design features that may be effective in metacognitive tutors. We emphasize, however, that the purpose of the discussion is to generate hypotheses, rather than to formulate definitive lessons learned.

The reader may recall the four levels of goals with respect to metacognitive tutoring, namely, to improve students' metacognitive behaviors, and their domain-level learning, both *during* the intervention and *afterwards*. Table 20.1 summarizes the results of our experiments with respect to the four goals. Three of the four studies show improved domain learning as a result of metacognitive support; two yield evidence that such support can improve metacognitive behavior during the intervention (see Table 20.1).

One might wonder whether metacognitive support that adapts to students' metacognitive behavior (i.e., what we have termed adaptive support), is more effective than static metacognitive support. Given that in our experience, adaptive metacognitive support is much more difficult to implement in computer tutors than static support, one might

Table 20.1 Relating the Four Studies to Four Goals of Metacognitive Support given in Figure 20.1

	<i>During intervention</i>		<i>After intervention</i>	
	<i>Improved metacognitive behavior</i>	<i>Improved domain learning</i>	<i>Improved metacognitive behavior</i>	<i>Improved domain learning</i>
Static	Cannot tell without online assessment of metacognitive behavior	Tutoring self-explanations leads to improvement	Did not assess	Did not assess
Adaptive	Scooter and the Help Tutor led to improved metacognitive behavior	Tutoring error correction and Scooter led to improvement; Help Tutor did not	Help Tutor did not improve metacognitive behavior in a transfer environment	Did not assess, but unlikely given lack of metacognitive improvement

wonder whether it provides a good return on investment. The pattern of results obtained does not provide support for the hypothesis that adaptive support leads to better domain-level learning than static support. The Self-Explanation Tutor, a form of static metacognitive support, improved students' domain-level learning, whereas the Help Tutor, with its detailed assessment of students' metacognition and highly specific metacognitive feedback, did not. However, we reiterate that in our experiments the systems differ in multiple ways, not just whether the support is static or adaptive, and that therefore it is hard to draw a firm conclusion. An ideal experiment would contrast this dimension, static versus adaptive support, for a single metacognitive skill, like self-explanation, in the same domain(s), like geometry.

We note further that our experiments do not at all rule out that adaptive metacognitive support may be better for long-term retention of students' metacognitive abilities (the third of the four goals in Figure 20.1). Our experiments are largely silent on this issue. We tested long-term improvement only with respect to help seeking, and did not find a significant improvement in students' help-seeking *behavior* (although we did find an improvement in their declarative knowledge of help seeking). It may well be, then, that across a broader range of metacognitive skills, adaptive metacognitive support will, upon further research, turn out to be desirable to better help students exhibit metacognitive behavior in the long term.

The pattern of results depicted in Table 20.1 seems to suggest another surprising trend: in the four systems we tested, the narrower the scope of metacognitive behaviors that the tutor targets (either through static or adaptive support), the stronger the effect on domain learning. Again, we note that our experiments were not designed to test this pattern in a rigorous manner; the systems differ along many dimensions in addition to this one. We found that the Self Explanation and Error Correction tutors, which each target a single metacognitive skill, were most clearly effective at the domain level. The Help Tutor on the other hand, which targets a wide range of metacognitive skills, with many detailed error categories corresponding to a comprehensive set of principles of help seeking, essentially did not have an effect on domain learning. Scooter falls somewhere in between these two "extremes," both in terms of the range of the metacognitive behaviors that it targets and its beneficial effect on domain-level learning. It may seem counter-intuitive that targeting a single significant metacognitive skill might be more beneficial (to domain learning) than providing feedback on a wide range of metacognitive behaviors. Perhaps students have

difficulty telling the forest from the trees when receiving a wide range of metacognitive feedback. Alternatively, systems that target a broad set of metacognitive skills may violate a key principle of both cognitive and metacognitive tutoring, that is, to communicate the goal structure underlying the problem solving (Roll, Alevan, McLaren, & Koedinger, 2007b).

The patterns of results shown in Table 20.1 further suggest that metacognitive support may be more effective when the students perceive the metacognitive steps that they are asked to perform as being a natural part of the task at hand and not as extra work. In the Self-Explanation and Error Correction tutors (which as mentioned were most clearly successful at the domain level), the targeted metacognitive behaviors were essential to successful completion of the task presented to the student. Students could not finish the geometry problems assigned to them without successfully explaining their steps, and could not complete the Excel training without successfully correcting their errors. This tight integration with the task makes students engage in the target metacognitive skill as a matter of course, and was shown to achieve the most improvement in learning.

This tight integration was not the case for Scooter and the Help Tutor, which were less successful at improving domain-level learning. Unlike the previous two examples, in the cases of Scooter and the Help Tutor, students' metacognitive errors did not prevent them from completing the task at hand. On the contrary, one might even argue that the given tasks could be solved more easily by gaming the system or abusing its help facilities—that is, desired metacognitive behavior may be viewed as being at odds with fast completion of the tasks as given (even if it is hypothesized to improve long-term learning). In particular, some students complained that Scooter prevented them from completing the problems.

Thus, with the caveat that there were multiple differences between the systems that might help explain the pattern of results in Table 20.1, in the systems that were more successful at the domain level, the metacognitive behaviors appeared to be a regular (natural?) part of task performance and so may have been perceived by students as being *beneficial* for the problem-solving process.

The Role of Affect in Metacognitive Choices

Another possible explanation for the differences in domain learning observed between studies is the extent to which the metacognitive process has connections with students' affect. It may be that the more connected the metacognitive process is to affective issues (e.g., gaming the system, and the gaming-related errors in help seeking may be more strongly related to affective issues than self-explanation or error correction), the less chance that an approach centered on scaffolding, monitoring, and tutoring will work. A source of suggestive evidence comes from the studies of gaming and affect that identified associations between negative affect and high levels of gaming (e.g., bored students are more likely to begin gaming). To confirm this hypothesis we will need corresponding evidence that self-explanation and error correction are not associated with negative affect. However, the association between gaming and affect already suggests that it may be valuable to either scaffold students' affect, in order to enable them to learn key metacognitive skills, or to assist students in developing affective self-regulation strategies that support better metacognition.

Tutoring and Assessing Metacognition During Learning

Beyond the direct effects on domain learning, the projects discussed in this chapter demonstrate the potential of intelligent tutoring systems to support the exercise of metacognitive

skills. Through a rigorous process of design and evaluation, using existing knowledge on domain-level tutoring, we created interventions that support students' metacognition within intelligent tutoring systems.

We have made some progress toward the goal of testing whether metacognitive support leads to measurable improvements in students' metacognition. The three adaptive systems presented in this chapter each depended on the development of sophisticated models that automatically monitor and assess differences in students' metacognitive behaviors, as they used the tutoring software. The Help Tutor used a rational production-rule model, as did the Error Correction tutor (although it was much simpler in its metacognitive components). Scooter used a statistical model developed through machine learning. These models have been one of the key innovations underpinning our work, enabling us to study students' metacognition more effectively by applying the models to existing data (e.g., to study how gaming varies across tutor lessons) as well as enabling automated interventions. An important future step will be to use these detectors to assess student metacognitive behavior after the intervention has been removed and the experimenters have returned to the laboratory. By assessing students' future metacognition, we will be in a position to test the hypothesis that adaptive metacognitive support will lead to better learning—and to find out how durable any effects are.

While much work is still to be done, we would like to emphasize two main lessons learned with respect to evaluation and assessment of metacognition:

- 1 Evaluation of interventions: Using the *in vivo* experimentation paradigm, we tested these interventions in ecologically valid classroom environments. *In vivo* experimentation is especially important in evaluating metacognitive instruction, which is tightly related to aspects of motivation, goal orientation, socio-cultural factors, etc. A metacognitive intervention that is effective in the laboratory may not be effective in more ecologically valid settings. Furthermore, evaluating the systems in students' "natural habitat" allows for collecting fine-grained data of students' metacognitive behaviors in the classroom.
- 2 Automated assessment of metacognitive behavior: Each of the four systems used automated assessment of students' actions at the domain level. Furthermore, three of the systems used models of students' metacognitive behavior in order to detect metacognitive errors, independently from domain errors in the cases of the Help Tutor and Scooter. Each of these models/detectors has been shown to be able to transfer, to at least some degree, across domain contents and student cohorts. This automation creates the potential that formative metacognitive assessment can be embedded within a wide variety of learning environments; a step that, when combined with appropriate instruction, may have the potential to achieve widespread improvements to student metacognition, with lasting effects on learning.

To summarize, the projects described above demonstrate advancements and possibilities in using intelligent tutoring systems to support, teach, and assess metacognition. At the same time, they suggest important links between affect and metacognition.

References

- Aleven, V. & Koedinger, K. R. (2000a). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000* (pp. 292–303). Berlin: Springer Verlag.
- Aleven, V. & Koedinger, K. R. (2000b). The need for tutorial dialog to support self-explanation.

- In C. P. Rose & R. Freedman (Eds.), *Building Dialogue Systems for Tutorial Applications, Papers of the 2000 AAAI Fall Symposium* (pp. 65–73). Technical Report FS-00-01. Menlo Park, CA: AAAI Press.
- Aleven, V. & Koedinger, K. R. (2002). An effective meta-cognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, 26(2), 147–179.
- Aleven, V., Koedinger, K. R., Sinclair, H. C., & Snyder, J. (1998). Combatting shallow learning in a tutor for geometry problem solving. In B. P. Goettl, H. M. Half, C. L. Redfield, & V. J. Shute (Eds.), *Intelligent Tutoring Systems, Fourth International Conference, ITS '98* (pp. 364–373). Lecture Notes in Computer Science 1452. Berlin: Springer Verlag.
- Aleven, V., McLaren, B., Roll, I., & Koedinger, K. R. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence and Education*, 16, 101–128.
- Aleven, V., Ogan, A., Popescu, O., Torrey, C., & Koedinger, K. R. (2004). Evaluating the effectiveness of a tutorial dialogue system for self-explanation. In J. C. Lester, R. M. Vicario, & F. Paraguaçu (Eds.), *Proceedings of Seventh International Conference on Intelligent Tutoring Systems, ITS 2004* (pp. 443–454). Berlin: Springer Verlag.
- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. M. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73(2), 277–320.
- Anderson, J. R. & Lebière, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.
- Arbreton, A. (1998). Student goal orientation and help-seeking strategy use. In S. A. Karabenick (Ed.), *Strategic help seeking: Implications for learning and teaching* (pp. 95–116). Mahwah, NJ: Lawrence Erlbaum Associates.
- Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S., & Woolf, B. P. (2007) Repairing disengagement with non-invasive interventions. *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, 195–202.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Combining fading with prompting fosters learning. *Journal of Educational Psychology*, 95, 774–783.
- Azevedo, R. & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, 96(3), 523–35.
- Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, 29, 344–70.
- Baker, R. S. J. d. (2007). Is gaming the system state-or-trait? Educational data mining through the multi-contextual application of a validated behavioral model. *Complete On-Line Proceedings of the Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling*, 76–80.
- Baker, R. S. J. d., Corbett, A. T., Koedinger, K. R., Evenson, S. E., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J., & Beck, J. (2006) Adapting to when students game an intelligent tutoring system. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 392–401.
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students “game the system”. *Proceedings of ACM CHI 2004: Computer-Human Interaction*, 383–390.
- Baker, R. S. J. d., Corbett, A. T., Roll, I., & Koedinger, K. R. (2008a) Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 18, 3, 287–314.
- Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008b). Why students engage in “gaming the system” behavior in interactive learning environments. *Journal of Interactive Learning Research*, 19(2), 185–224.
- Biswas, G., Leelawong, K., Schwartz, D., Vye, N. & The Teachable Agents Group at Vanderbilt (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19, 363–392.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In

- J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000) *How people learn: Brain, mind, experience, and school*. Washington, DC: National Research Council.
- Brown, A. (1987). Metacognition, executive control, self control, and other mysterious mechanisms. In F. Weinert & R. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Hillsdale, NJ: Erlbaum.
- Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In *Handbook of child psychology, Vol. III, Cognitive development* (pp. 77–166). New York: Wiley.
- Bruer, J. (1993). The mind's journey from novice to expert. *American Educator*, 6–46.
- Butler, D. L. (1998). The strategic content learning approach to promoting self-regulated learning: A report of three studies. *Journal of Educational Psychology*, 90(4), 682–697.
- Byrnes, J. P. & Wasik B. A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology*, 27, 777–786.
- Cheng, R., & Vassileva, J. (2005) Adaptive reward mechanism for sustainable online learning community. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, 152–159.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145–182.
- Chi, M. T. H., de Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439–477.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121–152.
- Conati, C., Muldner, K., & Carenini, G. (2006). From example studying to problem solving via tailored computer-based meta-cognitive scaffolding: Hypotheses and design. *Technology, Instruction, Cognition, and Learning*, 4(2), 139–190.
- Conati, C. & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, 11, 398–415.
- Conati, C. & VanLehn, K. (2001). Providing adaptive support to the understanding of instructional material. In C. Sidner & J. Moore (Eds.), *Proceedings of IUI 2001, the Sixth International Conference on Intelligent User Interfaces* (pp. 41–47). New York: ACM Press.
- Corbett, A. T. & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4, 253–278.
- Corbett, A. T. & Anderson, J. R. (2001) Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. *Proceedings of CHI 2002, Human Factors in Computing Systems* (pp. 245–252) (March 31–April 5, 2001, Seattle, WA, USA), ACM, 2001.
- Corbett, A. T., McLaughlin, M. S., Scarpinatto, K. C., & Hadley, W. S. (2000). Analyzing and Generating Mathematical Models: An Algebra II Cognitive Tutor Design Study. *Intelligent tutoring systems: Proceedings of the Fifth International Conference, ITS 2000* (pp. 314–323).
- Corbett, A. T., Wagner, A., Lesgold, S., Ulrich, H., & Stevens, S. (2006). The impact on learning of generating vs. selecting descriptions in analyzing algebra example solutions. *Proceedings of the 7th International Conference of the Learning Sciences* (pp. 99–105).
- Corbett, A. T., Wagner, A., & Raspat, J. (2003). The Impact of analyzing example solutions on problem solving in a pre-algebra tutor. *Proceedings of AIED 2003: The 11th International Conference on Artificial Intelligence and Education* (pp. 133–140).
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911.
- Fuson, K. C. (1990). Conceptual structures for multiunit numbers: Implications for learning and teaching multidigit addition, subtraction, and place value. *Cognition and Instruction*, 7, 343–403.
- Grosse, C. S. & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, 17(6), 612–634.

- Guthrie, J. T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated instruction on motivation and strategy use in reading. *Journal of Educational Psychology*, 92(2), 331–341.
- Hausmann, R. G. M., & Chi, M. T. H. (2002). Can a computer interface support self-explaining? *International Journal of Cognitive Technology*, 7(1), 4–14.
- Hiebert, J. & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction*, 14, 251–283.
- Karabenick, S. & Newman, R. (2006). *Help seeking in academic settings: Goals, groups, and contexts*. Mahwah, NJ: Erlbaum.
- Koedinger, K. R. & Alevan, V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review*, 19(3), 239–264.
- Koedinger, K. R. & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences*. Cambridge University Press.
- Kort, B., Reilly, R., & Picard, R. D. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy—building a Learning Companion. *Proceedings of IEEE Intl. Conf. on Advanced Learning Technologies*, 43–46.
- Magnussen, R. & Misfeldt, M. (2004). Player transformation of educational multiplayer games. *Proceedings of Other Players*. Retrieved from https://pure.dpu.dk/ws/fbspretrieve/82/player_transformation.pdf.
- Martin, B., Koedinger, K. R., Mitrovic, A., & Mathan, S. (2005). On using learning curves to evaluate ITS. *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, 419–428.
- Martínez Mirón, E. A., du Boulay, B., & Luckin, R. (2004) Goal achievement orientation in the design of an ILE. *Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments*, 72–78.
- Mathan, S. A. & Koedinger, K. R. (2005) Fostering the Intelligent Novice: Learning from errors with metacognitive tutoring. *Educational Psychologist*, 40(4), 257–265.
- Mevarech, Z. R. & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology*, 73, 449–71.
- Murray, R. C. & VanLehn, K. (2005) Effects of dissuading unnecessary help requests while providing proactive help. *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, 887–889.
- Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra word problem comprehension and its implications for the design of computer learning environments. *Cognition and Instruction*, 9(4), 329–389.
- Newman, R. S. (1994). Academic Help Seeking: a strategy of self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 293–301). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Newman, R. S. (2002). How self-regulated learners cope with academic difficulty: The role of adaptive help seeking. *Theory into Practice*, 41(2), 132–138.
- Paas, F. G. W. C. & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86, 122–133.
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117–175.
- Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31, 459–70.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385–407.
- Pintrich, P. R. & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40.
- Plöetzner, R., & VanLehn, K. (1997) The acquisition of qualitative physics knowledge during textbook-based physics training. *Cognition and Instruction*, 15 (2), 169–205.
- Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The

- effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, 23, 90–108.
- Rodrigo, M. M. T., Baker, R. S. J. d., d’Mello, S., Gonzalez, M. C. T., Lagud, M. C. V., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., Santillano, J. Q., Sugay, J. O., Tep, S., & Viehland, N. J. B. (2008). Comparing learners’ affect while using an intelligent tutoring system and a simulation problem solving game. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 40–49.
- Rodrigo, M. M. T., Baker, R. S. J. d., Lagud, M. C. V., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., Santillano, J. Q., Sevilla, L. R. S., Sugay, J. O., Tep, S., & Viehland, N. J. B. (2007). Affect and usage choices in simulation problem solving environments. *Proceedings of Artificial Intelligence in Education 2007*, 145–152.
- Roll, I., Alevén, V., McLaren, B. M., & Koedinger, K. R. (2007a). Can help seeking be tutored? Searching for the secret sauce of metacognitive tutoring. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education, AIED 2007* (pp. 203–210). Amsterdam, the Netherlands: IOS Press.
- Roll, I., Alevén, V., McLaren, B. M., & Koedinger, K. R. (2007b). Designing for metacognition—applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning*, 2(2), 125–140.
- Roll, I., Baker, R. S., Alevén, V., McLaren, B., & Koedinger, K. (2005). Modeling students’ metacognitive errors in two intelligent tutoring systems. In L. Ardissono, P. Brna, & A. Mitrovic (Eds.), *Proceedings of the 10th International Conference on User Modeling, UM 2005* (pp. 379–388). Berlin: Springer-Verlag.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479. Retrieved July 21, 2008, from Research Library Core database. (Document ID: 4466263).
- Scardamalia, M. & Bereiter, C. (1985). Development of dialectical processes in composition. In D. R. Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language, and learning: The nature and consequences of reading and writing*. Cambridge: Cambridge University Press.
- Schmidt, R. A. & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3 (4), 207–217.
- Schoenfeld, A. (1983). Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. *Cognitive Science*, 7, 329–363.
- Schoenfeld, A. H. (1987). What’s all the fuss about metacognition? In *Cognitive science and mathematics education* (pp. 189–215). Hillsdale, NJ: Erlbaum.
- Schofield, J. (1995). *Computers and classroom culture*. Cambridge, UK: Cambridge University Press.
- Shih, B., Koedinger, K. R., and Scheines, R. (2008). A response time model for bottom-out hints as worked examples. In R. S. J. d. Baker, T. Barnes, & J. E. Beck (Eds.), *Proceedings of the First International Conference on Educational Data Mining*, 117–126.
- Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Garnott & J. Parziale (Eds.), *Microdevelopment: A process-oriented perspective for studying development and learning* (pp. 31–58). Cambridge, MA: Cambridge University Press.
- Walonoski, J. A., Heffernan, N. T. (2006) Prevention of off-task gaming behavior in intelligent tutoring systems. In M. Ikeda, K. D. Ashley, & T. W. Chan (Eds.), *Proceedings of 8th International Conference on Intelligent Tutoring System, ITS 2006* (pp. 722–724). Berlin: Springer Verlag.
- Weerasinghe, A. & Mitrovic, A. (2005). Supporting self-explanation in an open-ended domain. In *Knowledge-based intelligent information and engineering systems* (pp. 306–313). Berlin: Springer Verlag.
- White, B. Y. & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction* 16 (1), 3–118.
- Wood, H. A. & Wood, D. J. (1999). Help seeking, learning, and contingent tutoring. *Computers and Education*, 33(2), 153–169.

Part X

Measurement

21 Measuring Metacognitive Judgments

Gregory Schraw

University of Nevada, Las Vegas

This chapter examines the measurement of metacognitive judgments and compares six different measures of absolute accuracy and relative accuracy. A lively debate has centered on how to measure metacognitive judgments over the last two decades (Allwood, Jonsson, & Granhag, 2005; Dunn, 2004; Juslin, Olsson, & Winman, 1996; Keren, 1991; Nelson, 1996; Schraw, 1995; Stewart, Brown, & Chater, 2005; Wright, 1996). I take the position that different outcome measures provide different types of information that complement one another. While some outcome measures are better suited for specific research questions, it is not the case that one measure is best in all situations. I propose that it is essential to understand different types of judgment measures and to use them appropriately to pose and answer useful research questions.

This chapter is divided into six sections. The first section provides a taxonomy of metacognitive judgments, including those made before, during, or after task performance. Section two describes three measures of absolute accuracy and three measures of relative accuracy that have been used in previous research. The third section provides a brief overview of some of the experimental treatments investigated in recent studies. These fall into three broad categories referred to as individual differences (e.g., working memory), task parameters (e.g., immediate versus delayed judgments), and text and test parameters (e.g., length or familiarity of text). The fourth section discusses several unresolved issues that affect the inferences drawn from monitoring research studies. Section five makes recommendations for future practice, while section six makes recommendations for future research.

A Taxonomy of Metacognitive Judgments

Many terms have been used to describe different aspects of metacognition since Flavell's seminal work to coin the terms metacognition and metamemory (Flavell, 1979; Flavell & Wellman, 1977). I use the superordinate terms metacomprehension, metamemory, metacognition, and metacognitive judgments to describe different types of cognitive activity. Metacomprehension refers to understanding at the broadest possible level that is necessary for an individual to be fully self-regulated (Pintrich, 2000; Schunk & Zimmerman, 2006; White & Frederiksen, 2005; Winne, 2001; Zimmerman, 2000). At least two components of metacomprehension are necessary for comprehensive understanding, including metamemory and metacognition. Metamemory refers to knowledge and understanding of memory in general, as well as one's own memory in particular (Nelson, Narens, & Dunlosky, 2004; Smith & Reio, 2006). This knowledge enables individuals to assess task demands on memory and to assess available knowledge and strategies in memory. Metacognition refers to knowledge about cognition and cognitive processes (Lin & Zabricky, 1998; McCormick, 2003). Metacognition often is partitioned

into declarative knowledge (i.e., knowing what), procedural knowledge (i.e., knowing how), and conditional knowledge (i.e., knowing when and why).

Metacognitive judgments refer to one of three different classes of judgments about one's learning and performance that are summarized in Table 21.1, which are labeled as prospective, concurrent, and retrospective judgments to indicate when the judgment is made with respect to the focal performance task. Prospective judgments (i.e., predictions) require the examinee to make a judgment about learning or performance prior to performing the criterion task. Concurrent judgments require the examinee to make confidence or performance judgments while performing the task. In a typical study, an individual may complete a 15-item test and make a confidence or performance accuracy judgment after each item. Retrospective judgments (i.e., post-dictions) require the examinee to judge the ease of learning or performance after completing a study phase or test. Retrospective judgments often are holistic in nature in which an examinee evaluates performance on all test items rather than an item-by-item basis as in concurrent judgments.

Three different types of prospective judgments are of interest to researchers, including judgments of learning, feeling of knowing, and ease of knowing. Judgments of learning (referred to in the literature as JOLs) have been researched extensively. The typical experimental format for assessing JOLs is for an examinee to study to-be-learned information such as a list of words, and then make predictions of subsequent recollection for each item. JOLs presumably tap metacognitive judgments about one's ability to encode and retain information. Feeling of knowing judgments (referred to in the literature as FOKs) occur when an individual is asked to predict whether they will *recognize* information that could not be *recalled* from long-term memory or from a prior study episode. FOKs assess one's ability to monitor the contents of memory and one's current ableness to search memory. Ease of learning judgments (referred to in the literature as EOLs) refer to judgments about the amount of time or effort needed to learn material in order to meet a subsequent recollection criterion. EOLs presumably measure one's ability to monitor the relative difficulty of comprehension processes.

Table 21.1 A Taxonomy of Calibration Judgments

<i>Time of judgment</i>	<i>Type of judgment</i>	<i>Description</i>
Prospective (judgments made prior to testing)	Judgments of Learning (JOL)	Judgments of subsequent recollection of recently studied information
	Ease of Learning (EOL) Judgments	Judgments prior to study about the relative ease of learning information
	Feeling of Knowing (FOK)	Judgments of subsequently recognizing information that could not be recalled
Concurrent (judgments made during testing)	Online Confidence Judgments	Judgments of confidence in one's performance
	Ease of Solution Judgments	
	Online Performance Accuracy Judgments (Calibration of performance)	Judgments about the accuracy of performance
Retrospective (judgments made after testing)	Ease of Learning/Solution	Judgments after study or testing about the relative ease of learning information
	Retrospective Judgments of Performance Accuracy	Judgments of how well one performed after completing all items

Concurrent judgments refer to ongoing assessments of learning or performance. These include confidence judgments (i.e., confidence about learning or performance), ease of solution, and performance accuracy judgments (i.e., calibration of performance) during an ongoing task. The typical experimental format is similar for all three measures, in which an individual answers a test item or performs a criterion task, and immediately makes a judgment regarding confidence, ease of problem solution, or performance accuracy. Concurrent judgments almost always are made on an item-by-item basis rather than over a set of items as is characteristic of retrospective judgments. Concurrent confidence judgments assess the individual's ability to gauge their performance as it occurs. Ease of solution judgments tap the ability to monitor task difficulty relative to available cognitive resources. Judgments of performance accuracy provide a measure of ability to monitor performance on the criterion task.

Retrospective judgments refer to judgments of learning or performance after the criterion task has been completed. These include item-by-item (i.e., local) as well as global judgments in which a single ease of learning or performance judgment is made for the entire criterion task. Retrospective ease of learning or ease of solution judgments are identical to those described above with the exception that they are made after all aspects of the criterion task have been completed. For example, after completion of a 40-item test, individuals may make retrospective ease of learning or solution judgments even if they have made item-by-item judgments previously. Retrospective performance accuracy judgments follow the same pattern.

Six Types of Metacognitive Judgment Scores

Most studies in the metacognitive monitoring literature focus on the relationship between confidence judgments and performance. At least six distinct types of outcome scores had been used to measure the goodness of fit of these judgments that can be classified as measures of absolute and relative accuracy. It is important for researchers to understand the distinction between absolute and relative accuracy. Indices of absolute accuracy measure whether a specific confidence judgment matches performance exactly, whereas indices of relative accuracy measure the relationship between correct and incorrect confidence judgments, or a set of confidence judgments and performance outcomes (Connor, Dunlosky, & Hertzog, 1997; Maki, Shields, Wheeler, & Zacchilli, 2005). Absolute accuracy provides a measure of the precision with which a person can judge test item performance. Measures of absolute accuracy include the absolute accuracy index, Hamann coefficient, and bias index. Relative accuracy describes the consistency of judgments, or how well a person can discriminate better-learned material from lesser-learned material. Measures of relative accuracy include the correlation coefficient, gamma coefficient, and discrimination index (Allwood et al., 2005; Keren, 1991; Nelson, 1996; Schraw, 1995; Yates, 1990). Table 21.2 provides a tabular classification and summary of each score's interpretation.

Measures of Absolute Accuracy

Absolute accuracy assesses the precision of a confidence judgment compared to performance on a criterion task. For example, if an individual is 100 percent confident and answers a test item correctly, we would conclude that the individual is highly accurate. If, in contrast, the individual is 50% confident, yet answers the test item correctly, we would conclude that the individual is moderately accurate. The squared deviation score for each item is then summed and divided by the total number of items to obtain the average absolute accuracy index.

Table 21.2 Types of Outcome Measures

<i>Construct being measured</i>	<i>Outcome measure</i>	<i>Score interpretation</i>
Absolute accuracy	Absolute accuracy index	Discrepancy between a confidence judgment and performance
	Hamann coefficient	Discrepancy between hits and misses in a contingency table
	Bias index	The degree of over- or under-confidence in judgments
Relative accuracy	Correlation coefficient	Relationship between set of judgments and corresponding performances scores
	Gamma coefficient	Dependence between judgments and performance.
	Discrimination index	Ability to discriminate between correct and incorrect outcomes

One way to measure absolute accuracy is using formula 1 below:

$$\text{Absolute accuracy index} = 1/n \sum_{i=1}^n (c_i - p_i)^2 \quad (1)$$

where c_i corresponds to a confidence rating for each item, p_i corresponds to a performance score for the same item, and n equals the number of confidence judgments. Most researchers ask examinees to rate confidence on a continuous scale ranging from 1 to 100%, or on an ordinal scale with 10-point intervals that range from 10% to 100% with intervals at 20%, 30%, etc. Similarly, performance is scored as 0% (i.e., incorrect) or 100% (i.e., correct). Each deviation score between confidence and performance is squared to place it on a positive scale that ranges from zero to the upper positive end of the scale being employed. Scores close to zero correspond to high accuracy, while large positive scores toward the upper end of the scale correspond to low accuracy.

Absolute accuracy as shown in formula 1 measures the discrepancy between a confidence judgment and performance by computing the squared deviation between them on the same scale. Smaller deviations correspond to better accuracy. Formula 1 provides a measure of “absolute” accuracy to the extent that a person’s confidence judgment is compared in an absolute fashion to their performance on the same task.

Formula 1 is appropriate when continuous data are used for confidence and performance. An equivalent formula is given by Allwood et al. (2005) for categorical data. The Hamann coefficient shown in formula 2 is appropriate when using categorical data in a 2×2 data array as shown in Table 21.3 (Agresti, 1984, 1990; Romesburg, 1984; Schraw, 1995; Nietfeld, Enders, & Schraw, 2006; Siegel & Castellan, 1988).

$$\text{Hamann coefficient} = \frac{(a + d) - (b + c)}{a + d + b + c} \quad (2)$$

Table 21.3 includes four types of outcomes, including two hits (i.e., cells a and d) and

Table 21.3 A 2x2 Performance–Judgment Data Array

Judgment	Performance		Row
	Yes	No	marginals
Yes	a (Hit)	b (Miss)	a + b
No	c (Miss)	d (Hit)	c + d
Column Marginals	a + c	b + d	a + b + c + d

two misses (i.e., cells *b* and *c*). Formula 2 shows that the Hamann coefficient provides the number of hits minus the number of misses over all possible outcomes. Thus, it is conceptually similar to the discrepancy model that underlies formula 1.

Formulae 1 and 2 provide measures of absolute accuracy that focus on the discrepancy between confidence judgments and performance for an individual (Schraw, 1995). Each discrepancy is item-specific; thus, it provides an index of the degree of precision for each item using the examinee’s performance on that item as the point of absolute comparison. The sum of accuracy scores provides a measure of composite accuracy for each individual.

Another measure of absolute accuracy is the bias index, which assesses the degree to which an individual is over- or underconfident when making a confidence judgment. Bias may be evaluated using formula 3 shown below.

$$\text{Bias index} = 1/n \sum_{i=1}^n (c_i - p_i) \tag{3}$$

where *c_i* corresponds to a confidence rating and *p_i* corresponds to a performance score. Formula 3 differs from formula 1 only in that the discrepancy between confidence and performance is not squared. This allows the deviation between the two to be signed positively or negatively. Thus, the bias index provides information about the direction and magnitude of the lack of fit between confidence and performance. When confidence is high and performance is low, overconfidence occurs. When confidence is low and performance is high, underconfidence occurs. Like the absolute accuracy index, the bias index focuses on the disparity between a specific confidence judgment and performance outcome, which provides a measure of bias that is absolute rather than relative.

Measures of Relative Accuracy

Relative accuracy assesses the relationship between confidence judgments and performance scores on a criterion task. Relative accuracy typically is assessed using some type of correlational measure such as Pearson’s *r* or a contingency coefficient such as gamma (Nelson, 1996). Measures of relative accuracy focus on the consistency of a set of confidence judgments relative to a set of performance outcomes rather than on the degree to which each confidence judgment is precise on an item-to-item basis.

The formula for r using standardized variables is shown in formula 4:

$$\text{Pearson correlation coefficient} = \Sigma Z_x Z_y / (n - 1) \quad (4)$$

where $Z_x = \Sigma (X_i - \bar{X})/s$.

It is important to note that formulae 1 and 3 focus on the discrepancy between a confidence judgment and performance outcome, whereas formula 4 focuses on the relationship between two different types of discrepancies; the difference between a confidence judgment and mean confidence, and the difference between performance on an item and mean performance. Formula 4 provides information about confidence judgments relative to performance outcomes. A positive relationship indicates that deviation scores are in the same direction. A negative relationship indicates that deviation scores are in the opposite direction. Thus, while absolute accuracy focuses on the precision of a specific judgment compared to a performance outcome, relative accuracy focuses on the trend in confidence judgments relative to the trend in performance outcomes. An individual who makes consistent confidence judgments may be high with respect to relative accuracy without necessarily being high with respect to absolute accuracy (Juslin et al., 1996).

Gamma (Goodman & Kruskal, 1954) provides a measure of relative accuracy that is appropriate when using categorical data in a 2×2 data array as shown in Table 21.3 (Nelson, 1996; Schraw, 1995).

$$\text{Gamma coefficient} = \frac{(a)(d) - (b)(c)}{(a)(d) + (b)(c)} \quad (5)$$

Although similar, the Hamann and gamma coefficients differ in several important ways (Nietfeld et al., 2006; Schraw, 1995), the most important of these being that the Hamann coefficient is based on discrepancies among the additive values within each of the four cells shown in Table 21.3, whereas Gamma is based on multiplicative values. These differences lead to corresponding differences in the interpretation of the measures (Schraw, 1995). Nelson and colleagues (Nelson, 1984; Nelson & Gonzalez, 1996) have discussed in detail situations in which it is appropriate to use gamma.

Discrimination assesses the degree to which an individual distinguishes between confidence judgments for correct versus incorrect items. Variations on the discrimination index have been reported by Allwood et al. (2005) (i.e., the resolution index) and by Glenberg and Epstein (1987) (i.e., the confidence-judgment accuracy quotient), which are related to d' measures used in signal detection theory (Keren, 1991; Yates, 1990). A general discrimination formula is shown below:

$$\text{Discrimination index} = 1/n \sum_{i=1}^{n_c} (c_{i,\text{correct}}) - 1/n \sum_{i=1}^{n_i} (c_{i,\text{incorrect}}) \quad (6)$$

where c_{correct} corresponds to confidence for each correct item, $c_{\text{incorrect}}$ corresponds to confidence for each incorrect item, n_c to the total number of items answered correctly, and n_i to the total number of items answered incorrectly. Formula 6 provides useful information about an individual's ability to distinguish between confidence for correct and correct items. Positive discrimination occurs when an individual is more confident about correct versus incorrect items, whereas negative discrimination occurs when an individual is more

confident about incorrect items. The discrimination index differs from the absolute accuracy and bias indices in that it is computed on aggregated data, rather than individual items that subsequently are aggregated. Sprenger and Dougherty (2006) have suggested that it provides a relative measure. The discrimination index captures a facet of metacognitive monitoring that is not captured by absolute accuracy and bias measures; that is, to monitor one's confidence across items rather than to monitor one's confidence about one's performance on a specific item.

Hybrid Scores

Several researchers have proposed hybrid indices that combine components of accuracy, bias, and discrimination all within the same measure (Keren, 1991; Yates, 1990) or alternative hybrids based on signal detection theory (Dunn, 2004; Stewart et al., 2005). For example, Keren (1991) discussed the Brier Score, which includes both an absolute accuracy and a discrimination component. Presently, it is unknown whether hybrid scores are more useful because they include more information. One potential problem is that it is difficult to interpret composite measures in a succinct manner. Studies that have used multiple measures typically report and interpret each measure individually, while also attempting to synthesize all measures into a holistic interpretation (Allwood et al., 2005; Maki et al. 2005). It seems prudent given the above discussion to report individual outcome measures such as absolute accuracy and discrimination even in situations where a hybrid score is used.

Keren (1991, pp. 5–6) also discussed the use of the calibration curve, which provides a plot of confidence judgments relative to performance scores. Plots of items on a test can be compared to an ideal calibration line with a slope of 1.0 in which items answered correctly 60% of the time receive an average confidence judgment of 0.60. Although calibration curves are not computed using a formula, it is possible to examine the deviation between plotted calibration scores and ideal calibration. The sum of these deviations would provide a measure of the discrepancy between observed calibration and ideal calibration. This information could be used to augment the information provided by formulas 1 and 3.

Summary of Outcome Scores

There are at least six possible measures that are suitable for assessing metacognitive judgments, including the absolute accuracy index, Hamann coefficient, bias index, correlation coefficient, gamma coefficient, and discrimination index. These measures have different interpretations that provide information about different facets of metacognitive monitoring. Measures of absolute accuracy provide indices of the discrepancy between a confidence judgment and performance on a task. They provide valuable information about the precision and direction of confidence judgments as well as the magnitude and direction of judgment error. Measures of relative accuracy provide indices of the relationship between a set of confidence judgments and a corresponding set of performance scores, or between correct and incorrect confidence judgments. They provide information about the relationship between a set of judgment and a set of performance outcomes.

An Overview of Existing Research

Researchers have investigated many different aspects of metacognitive judgments (i.e., prospective, concurrent, and retrospective) using a variety of independent variables, which can be classified into three categories referred to as individual difference, task parameter,

and text and test parameter variables. A comprehensive review lies beyond the scope of this chapter; however, I provide examples to illustrate the scope of recent research, as well as a brief summary of general findings.

Individual Differences

Individual differences refer to any attribute that an individual brings to the experimental environment. Individuals differ on many dimensions, but a relatively small number of variables have been of interest to researchers. One is the role that working memory plays in judgment accuracy. Dougherty and Hunt (2003) and Sprenger and Dougherty (2006) found that working memory capacity as measured by a span task was related to probability judgments, but not frequency judgments. This difference was attributed to the greater complexity of probability judgments, which required considerably more cognitive processing to reach a decision. In a related study, Souchay, Isingrini, Clarys, and Eustache, (2004) found that the relative accuracy of FOKs were correlated with executive cognitive functions, whereas JOLs were not.

A second variable of interest is general verbal ability. Gillstrom and Ronnberg (1995) found that students with better reading skills were better able to predict and judge performance on a reading task. Maki et al. (2005) reported that reading ability was related to absolute, but not relative accuracy using the bias index and gamma coefficient. Poor readers were more overconfident than good readers; however, correlations between judgments and performance did not differ between the groups.

A third critical variable is cultural differences, which appears to affect monitoring accuracy, in part, due to differences in overconfidence. Lundeberg, Fox, Brown, and Elbedour (2000) found that cultural differences played a large role in confidence differences, whereas gender played little role. These results were replicated by Wallsten and Gu (2003), who found that Chinese students were significantly more confident (and overconfident) than American or Japanese students.

Task Parameters

Task parameters refer to an attribute of the task that affects accuracy. Several variables have been studied in depth, including differences between immediate versus delayed metacognitive judgments, different types of judgment tasks, and strategy training. Metacognitive judgments typically are more accurate when they are made following a delay (Kimball & Metcalfe, 2003; Nelson et al., 2004; Thiede, Anderson, & Therriault, 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005). The delayed judgment effect presumably occurs due to additional opportunity to engage in metacognitive monitoring between the study and test phases of an experiment.

Different types of judgments affect performance on recall and recognition tests, and are more sensitive to materials and verbal skills (Maki et al., 2005). Kelemen, Frost and Weaver (2000) found weak relationships among four monitoring tasks, including ease of learning, feeling of knowing, judgments of learning, and text comprehension monitoring. These studies suggested that different metacognitive judgment tasks require different underlying metacognitive processes.

Use of strategies and strategy training also are related to judgment accuracy. Thiede and Anderson (2003) found that summarizing texts improved the relationship between confidence and performance. Son (2004) reported that distributed study improved relative accuracy. In related studies, Nietfeld, Cao and Osborne (2005) and Bol, Hacker, O'Shea and Allen (2005) found that monitoring accuracy remained stable overall the course of a

one-semester course even with ongoing monitoring practice. However, Koriat, Sheffer, and Ma'ayan (2002) reported that overconfidence may increase with practice. Of special interest is that none of these studies found that accuracy improved as a function of practice.

Text and Test Parameters

Text and test parameters refer to aspects of the text (i.e., length and complexity) and test items (e.g., difficulty) that affect performance or accuracy. A number of studies suggest that text and test factors have a significant effect on accuracy (Maki & Berry, 1984; Maki & Serra, 1992; Maki & Swett, 1987). For example, Lin, Zabrocky, and Moore (2002) found that text difficulty was related to relative accuracy using the gamma coefficient. Accuracy was maximized when passages were at the reader's current reading skill level rather than being too easy or too difficult. Weaver (1990) reported that the number of test items associated with a reading passage, rather than the number of reading passages, played a major role in score reliability and relative accuracy using the gamma coefficient. Weaver (1990) recommended that four to six items be used to assess judgment accuracy on each reading passage.

Burson, Larrick and Klayman (2006) found that test item difficulty was a significant constraint on performance accuracy judgments that affected both skilled and unskilled performers on the task. These findings replicated Schraw and Roedel (1994), who found that difficult items led to significantly more overconfidence than easy items using the bias index. Juslin, Wennerholm, and Olsson (1999) also found that test item format affected relative accuracy, but especially the degree to which individuals were overconfident. Collectively, these studies suggested that items should cover a range of difficulty with performance accuracy ranging from 40% to 75%.

Brewer, Sampiano, and Barlow (2005) reported that the type of information in a text affects relative accuracy using the gamma coefficient. Deceptive information leads to poorer accuracy than non-deceptive information. Similarly, information that is judged to be relevant to a judgment task improves accuracy, whereas irrelevant information leads to poorer discrimination (Dougherty & Franco-Watkins, 2003; Dougherty & Sprenger, 2006).

Summary

These studies collectively suggest that individual differences, task constraints, and test and text parameters affect accuracy. It is unclear whether differences observed in these studies can be attributed to a change in metacognitive ability, or to the facilitative effect of external scaffolding such as more time to make judgments, longer texts, or more test questions. Nevertheless, these studies make two important points. The first is that a variety of factors affect judgment accuracy, including aspects of the person, to-be-learned materials, and task. The second is that judgment accuracy varies within-person as a function of the type of judgment that is made and the type of outcome measure that is used to evaluate that judgment. In the next section, I focus on a variety of outcome measures that have been used to evaluate metacognitive monitoring.

Unresolved Issues

A number of unresolved issues remain regarding the six metacognitive outcome scores described above. However, two issues are of special importance. One is the extent to which

the measure is reliable. A second issue concerns the empirical relationship among different outcome scores.

Reliability

Previous research suggests that metacognitive judgments are not always measured in a reliable manner (Thompson, 1999; Thompson & Mason, 1996). Reliability is necessary, but not sufficient for validity; thus, it is essential that measures be reliable (Osterlind, 2006). There are at least three reasons why scores may have low or unacceptable reliability; that is, values less than 0.70. One is the number of judgments that the score is based on. Weaver (1990) showed that scores are far more reliable in a study of text comprehension when four to six test items are based on the same passage, rather than many passages that use only one test item. Indeed, testing theory suggests that reliability will be maximized as the number of items increases, when items have a sufficient range of difficulty, and when items are neither too easy nor too difficult (Osterlind, 2006).

A second reason for inadequate reliability is the time of assessment. A number of studies have investigated the delayed judgment of learning effect and concluded that delayed judgments are significantly more accurate than immediate judgments. Part of this effect undoubtedly is attributed to cognitive factors that affect judgment accuracy. Nevertheless, it may be the case that judgments made immediately are less reliable due to restricted range (i.e., more items are answered incorrectly) or other testing factors that affect reliability. This suggests that more items are needed at immediate testing to achieve adequate reliability.

A third reason pertains to the “grain size” of the metacognitive judgment. Currently, most studies investigate item-by-item judgments. However, a number of studies also investigate composite judgments in which an examinee answers 10 items, then judges overall performance on these items (e.g., Dunlosky, Rawson, & Middleton, 2005; Nietfeld et al., 2005). Retrospective judgments are especially likely to include judgments aggregated over multiple test items. Composite judgments pose two distinct problems. One is that only a single judgment is made; thus, it is impossible to calculate the reliability of that judgment. A second problem is that a composite judgment is based on multiple data points (i.e., the total number of problems answered), which permits smoothing of one’s estimate to match one’s sense of average performance. Unfortunately, there are no studies in the metacognitive monitoring literature that attempt to compare the reliability and utility of item-by-item versus composite judgments. This is an important topic for future research.

Relationships Between Outcome Scores

Virtually nothing is known about the relationships between the different types of outcome scores summarized in Table 21.2. Most studies that have used two or more of the outcomes scores in Table 21.2 report low inter-correlations, inconsistent patterns of results, or dissociations (Allwood et al., 2005; Conor et al., 1997; Dougherty & Sprenger, 2006; Kelemen et al., 2000; Maki et al., 2005; Nietfeld et al., 2006). This suggests that different outcome measures are unrelated in part because they assess different underlying aspects of metacognitive monitoring. The fact that different outcomes assess different constructs has been discussed previously. Schraw (1995) argued that the Hamann coefficient measures absolute accuracy while gamma measures relative accuracy. A follow-up Monte Carlo study revealed that the two measures had important differences with respect to distributional properties and predictive capability. Recent studies (e.g., Maki et al., 2005) clearly distinguish between measures of absolute and relative accuracy. Juslin et al. (1996)

also noted that correlational measures of relative accuracy such as gamma and Pearson's r fail to account for over- or underconfidence that is more appropriately measured by the bias index.

There are two extremely important implications regarding the paucity of knowledge about the relationships among outcome measures. One is practical in nature; it is unclear to researchers what measure to use in what particular circumstance. For this reason, researchers frequently use one or two measures out of methodological convention rather than strategic awareness. I offer some guidelines to ameliorate this confusion in the following section. A second implication is that researchers have failed to develop theoretical models that distinguish among the underlying metacognitive processes that affect judgments. Table 21.2 suggests a clear distinction among six different types of outcomes: absolute accuracy, the Hamann coefficient, bias, the correlation coefficient, gamma, and discrimination. While this probably is not an exhaustive list, it illustrates the need to construct models that explain and predict the underlying cognitive construct that is assessed by each of these measures. In addition, there is a pressing need for technical studies such as Nietfeld et al. (2006) that compare the distributional and predictive abilities of different outcome measures.

Recommendations for Future Practice

I make three general suggestions for best practice: (1) understand the construct being measured, (2) select the outcome that matches the construct being studied, and (3) use multiple outcome measures whenever possible. The first recommendation is to understand the construct being measured by each outcome measure in Table 21.2. To reiterate the main points above, absolute accuracy reveals the precision of a judgment with regard to actual performance. Absolute accuracy scores typically are computed using the deviation between a confidence judgment and performance on a test item. Scores close to a scaled value of 1 correspond to high accuracy, whereas scores close to zero correspond to low accuracy. This outcome measure should be used when researchers are interested in investigating whether a treatment enhances the goodness of fit between a confidence judgment and corresponding performance. Absolute accuracy is an ideal measure when implementing a treatment such as monitoring practice (Bol et al., 2005; Nietfeld et al., 2005) or reading strategies (Gillstrom & Ronnberg, 1995) that are intended to improve performance, monitoring accuracy, and especially the goodness of fit between performance and confidence or accuracy judgments about one's performance.

Relative accuracy typically is measured using an index of association such as gamma, the point-biserial correlation, or Pearson's r , and provides an index of the relationship between a set of judgments and corresponding performances scores. It is important to note that relative accuracy should be interpreted like a correlation; that is, it measures whether there is a correspondence or consistency between one set of scores and a second set of scores. It does not measure absolute precision, over- or underconfidence, or discrimination (Juslin et al., 1996). Indeed, as Schraw (1995) has noted, relative accuracy can be quite high despite low precision in an absolute sense, and a number of studies have reported dissociations between absolute and relative accuracy (Dougherty & Sprenger, 2006; Maki et al., 2005). The relative accuracy index should be used in studies that investigate whether a treatment intervention improves the consistency of judgment across a set of performance items.

The bias index provides a measure of over- or underconfidence in judgment, both in terms of direction and magnitude. Bias indicates the direction and magnitude of over- or underestimation in a judgment. This index should be used when researchers are interested

in investigating whether a treatment affects confidence judgments, and especially whether the treatment decreases or increases confidence relative to performance. As an example, bias would be an ideal indicator when comparing cross-cultural differences in confidence like those examined in Wallsten and Gu (2003).

Discrimination provides a measure of an individual's ability to discriminate between correct and incorrect outcomes, and therefore can be used to determine whether confidence for correct outcomes is greater than confidence for incorrect outcomes. The discrimination provides information about the direction and magnitude of judgment errors when comparing confidence rating for correct versus incorrect responses. The discrimination index should be used when a researcher is interested in determining whether an individual can distinguish between correct versus incorrect performance.

A second recommendation is to select the outcome measure that is most closely aligned to the goals and hypotheses that drive the research. In essence, the six outcome measures in Table 21.2 assess the precision of judgments (i.e., absolute accuracy), the correspondence between judgments versus performance (i.e., relative accuracy), the direction and magnitude of judgment error (i.e., bias), and one's ability to discriminate between judgments for accurate versus inaccurate performance (i.e., discrimination). A researcher should choose one or more measures that are relevant to these goals.

A third recommendation is to use multiple measures whenever possible. As mentioned above, some outcome measures have become popular through conventional usage even when they are not the best measurement choices. There is no compelling reason to believe that the six types of measures described in this chapter are incompatible with one another. Indeed, it is reasonable to assume that these measures complement each other in that they provide overlapping, but not substantially redundant, types of information. I recommend that researchers report multiple outcomes whenever possible. This would inform one's findings to the fullest extent possible, as well as foster an emerging pool of data that could be used in future meta-analyses to address broader comparative questions about the appropriateness and distributional suitability of each measure.

Recommendations for Future Research

This chapter has focused on understanding and using different metacognitive outcome measures in applied settings. However, I wish to make several brief suggestions for future research, especially regarding the conceptual and technical issues of the six measures shown in Table 21.2. One pressing need is to develop theoretical models that distinguish among the underlying constructs involved in metacognitive judgments. Ideally, a model would describe the processes and resources required of different metacognitive judgments, as well as the cognitive and mathematical relationships among these constructs (Dunn, 2004; Nelson et al., 2004; Stewart et al., 2005).

A second direction is for additional technical validation work on each of the measures, but especially their inter-relationships. Nietfeld et al. (2006) conducted computer simulation studies that compared the Hamann and gamma coefficients. These studies provided a basic understanding of the distributional properties of each statistic, as well as estimation accuracy. Additional Monte Carlo studies are needed to compare the four measures described in formulae 1, 3, 4, and 6. These studies also may choose to compare Bayesian measures described by Juslin et al. (1996) and Dunn et al. (2004) to six outcome measures in Table 21.2. Previous research indicates that two or more measures are sometimes uncorrelated, which may be due in part to poor monitoring by participants in these studies. However, there are situations in which measures should be correlated, but especially when monitoring is highly accurate. In this case, both absolute and relative accuracy

should be close to perfect (i.e., an absolute accuracy score of 1.0, bias score of 0.0, and a correlation of 1.0), and discrimination should be perfect as well (i.e., a score of 1.0).

A third direction is to investigate the predictive validity of each measure. Currently, it is unknown whether the outcome scores in Table 21.2 explain the same amount of variation in distal outcomes such as learning and achievement in college courses. Similarly, it is unknown whether two different outcomes such as absolute accuracy and bias account for the same pool of variation in an outcome. Current findings from studies that use multiple outcome measures do not support a common variance model; however, future studies should be devoted to question and further explore the relative statistical and predictive independence of the scores.

Summary

This chapter has reviewed six types of outcome measures used in the metacognitive monitoring literature. Each measure assesses a somewhat different aspect of monitoring. All of these measures are complementary and therefore should be reported whenever possible. I have argued that little is known about the conceptual and technical aspects of each measure. Future research should focus on the development of a theory that explains the relationships between the constructs associated with each measure. In addition, technical studies are needed that compare the distributional properties of the six measures, their inter-correlations, as well as their predictive validity in applied settings.

References

- Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: John Wiley & Sons.
- Agresti, A. (1990). *Categorical data analysis*. New York: John Wiley & Sons.
- Allwood, C. M., Jonsson, A. C., & Granhag, P. A. (2005). The effects of source and type of feedback on child witnesses' metamemory accuracy. *Applied Cognitive Psychology, 19*, 331–344.
- Bol, L., Hacker, D., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *Journal of Experimental Education, 73*, 269–290.
- Brewer, W. F., Sampiano, C., & Barlow, M. R. (2005). Confidence and accuracy in recall of deceptive and non-deceptive sentences. *Journal of Memory and Language, 52*, 618–627.
- Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: Perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90*, 60–77.
- Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12*, 50–71.
- Dougherty, M. R. P. & Franco-Watkins, A. M. (2003). Reducing bias in frequency judgment by improving source monitoring. *Acta Psychologica, 113*, 23–44.
- Dougherty, M. R. P. & Hunt, J. E. (2003). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica, 113*, 263–282.
- Dougherty, M. R. P. & Sprenger, A. (2006). The influence of improper sets of information of judgment: How irrelevant information can bias judged probability. *Journal of Experimental Psychology: General, 135*, 262–281.
- Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypothesis. *Journal of Memory and Language, 52*, 551–565.
- Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*, 524–542.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist, 34*, 906–911.
- Flavell, J. H. & Wellman, H. (1977). Metamemory. In R. V. Vail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition*. Hillsdale, NJ: Lawrence Erlbaum.

- Gillstrom, A. & Ronnberg, J. (1995). Comprehension calibration and recall prediction accuracy of texts: Reading skill, reading strategies, and effort. *Journal of Educational Psychology, 87*, 545–558.
- Goodman, L. A. & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association, 49*, 732–764.
- Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness testimony: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 1304–1316.
- Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 1038–1052.
- Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition, 28*, 92–107.
- Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77*, 217–273.
- Kimball, D. R. & Metcalfe, J. (2003). Delayed judgments of learning affects memory, not metamemory. *Memory & Cognition, 31*, 918–929.
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased overconfidence with practice. *Journal of Experimental Psychology: General, 131*, 147–162.
- Lin, L. M. & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology, 23*, 345–391.
- Lin, L. M., Zabrucky, K. M., & Moore, D. (2002). Effects of text difficulty and adults' age on relative calibration of comprehension. *The American Journal of Psychology, 115*, 187–198.
- Lundeberg, M. A., Fox, P. W., Brown, A. C., & Elbedour, S. (2000). Cultural influences on confidence: Country and gender. *Journal of Educational Psychology, 90*, 152–159.
- Maki, R. H. & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory & Cognition, 10*, 663–679.
- Maki, R. H. & Serra, M. (1992). The basis of test predictions for text materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 116–126.
- Maki, R. H., Shields, M., Wheeler, A. E., Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology, 97*, 723–731.
- Maki, R. H. & Swett, S. (1987). Metamemory for narrative text. *Memory & Cognition, 15*, 72–83.
- McCormick, C. B. (2003). Metacognition and learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (pp. 79–102). Hoboken, NJ: John Wiley & Sons.
- Nelson, T. O. (1984). A comparison of current measures of feeling-of-knowing accuracy. *Psychological Bulletin, 95*, 109–133.
- Nelson, T. O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not the absolute performance on an individual item: Comments on Schraw (1995). *Applied Cognitive Psychology, 10*, 257–260.
- Nelson, T. O. & Gonzales, R. (1996). Measuring ordinal association in situations that contain tied scores. *Psychological Bulletin, 119*, 159–165.
- Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research in metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods, 9*, 53–69.
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *Journal of Experimental Education, 74*, 7–28.
- Nietfeld, J. L., Enders, C. K., & Schraw, G. (2006). A monte carlo comparison of measures of relative and absolute monitoring accuracy. *Educational and Psychological Measurement, 66*, 258–271.
- Osterlind, S. J. (2006) *Modern measurement: Theory, practice, and applications of mental appraisal*. Upper Saddle River, NJ: Prentice Hall.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts,

- P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.
- Romesburg, H. C. (1984). *Cluster analysis for researchers*. London: Wadsworth, Inc.
- Schraw, G. (1995). Measures of feeling-of-knowing accuracy: A new look at an old problem. *Applied Cognitive Psychology*, 9, 321–322.
- Schraw, G. & Roedel, T. D. (1994). Test difficulty and judgment bias. *Memory & Cognition*, 24, 63–69.
- Schunk, D. H. & Zimmerman, B. J. (2006). Competence and control beliefs: Distinguishing the means and the ends. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology*, 2nd ed. (pp. 349–367). San Diego: Academic Press.
- Siegel, S. & Castellan, N. J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- Smith, M. C. & Reio, T. G. (2006). Adult development, schooling, and the transition to work. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology*, 2nd ed. (pp. 115–138). San Diego: Academic Press.
- Son, L. K. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 601–604.
- Souchay, C., Isingrini, M., Clarys, D., & Eustache, F. (2004). Executive functioning and judgment-of-learning versus feeling-of-knowing in older adults. *Experimental Aging Research*, 30, 47–62.
- Sprenger, A. & Dougherty, M. R. (2006). Differences between probability and frequency judgments: The role of individual differences in working memory capacity. *Organizational Behavior and Human Decision Processes*, 99, 202–211.
- Stewart, N., Brown, G. D., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, 112, 881–911.
- Thiede, K. W. & Anderson, M. C. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, 28, 129–160.
- Thiede, K. W., Anderson, M. C., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95, 66–73.
- Thiede, K. W., Dunlosky, J., Griffin, D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 1267–1280.
- Thompson, W. B. (1999). Individual differences in memory monitoring accuracy. *Learning and Individual Differences*, 11, 365–376.
- Thompson, W. B. & Mason, S. E. (1996). Instability of individual differences in the association between confidence judgments and memory performance. *Memory & Cognition*, 94, 226–234.
- Wallsten, T. S. & Gu, H. (2003). Distinguishing choice and subjective probability estimation processes: Implications for theories of judgment and for cross-cultural comparisons. *Organizational Behavior and Decision Processes*, 90, 111–123.
- Weaver, C. A. (1990). Constraining factors in the calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 214–222.
- White, B. & Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist*, 40, 211–223.
- Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Mahwah, NJ: Erlbaum.
- Wright, D. B. (1996). Measuring feeling of knowing: Comment on Schraw (1995). *Applied Cognitive Psychology*, 10, 261–268.
- Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice-Hall.
- Zimmerman, B. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.

22 Sins Committed in the Name of Ecological Validity

A Call for Representative Design in Education Science

John Dunlosky
Kent State University

Sara Bottioli
University of Pavia

Marissa Hartwig
Kent State University

Do you want conclusions and findings from education research to generalize to real-world education, where students are taught, learn, and are tested? In critiquing education research, are you concerned when it is not ecologically valid? Most of us would answer “yes” to both of these questions. Even if we believe that broad generalizability may not be feasible in any one experiment, conducting “ecologically valid” research is no doubt an ideal to which we still strive. In the present chapter, we agree that generalizability is essential, but we deconstruct ecological validity and argue that the term is often misused and misrepresented in a manner that may ultimately stifle the development of methods for achieving generalizability. For those of you who answered both questions with a resounding “yes,” and the present authors certainly have done so, we must emphasize that our goal is not to undermine the ideal of validity—every social scientist’s desire is to conduct research that is generalizable to a target ecology. Indeed, we mean to hold this ideal with the highest regard, and by doing so, we will reveal the sins committed in the name of ecological validity and the promise of research programs that are informed by representative design.

By no means are we the first to call for this reform within the scientific community—to scrutinize the use of the term ecological validity and to attend more deeply to issues pertaining to generalizability. Others who have made such a call include Araújo, Davids, and Passos (2007), Hammond (1998), Petrinovich (1989), and Schmuckler (2001). Although each of these authors offer unique insights about conducting research that will generalize, they all converge on a common theme: The concept of ecological validity is inadequate for guiding the design of research that will generalize to target ecologies. Given that these papers are more extensive than the current chapter, we highly recommend reading them to get a more thoroughly informed and well-rounded perspective on ecological validity and representative design.

Here, we have the more modest goal of providing an introduction to these concepts. We first describe some of the sins committed in the name of ecological validity—partly what makes the concept less than ideal for understanding how to achieve generalizability. We then reveal the challenge of generalizability in more depth, by discussing just a subset of the numerous factors in the target ecology of education researchers. We then introduce the

work of Egon Brunswik, who originally defined ecological validity and argued that psychological scientists should use representative designs to increase the chances of obtaining generalizability.

Sins Committed in the Name of Ecological Validity

Although we intentionally included “sins” in the title of our chapter to be provocative, in a literal sense the use of the term ecological validity can lead us to commit offences that may sidetrack the development of scientific research intended to improve student scholarship. These sins arise when we mistakenly reject research as not being ecologically valid, and they arise when we mistakenly accept research as being ecologically valid. Even we have used the term inappropriately in the past. Thus, our goal is not to cast the first stone against others, but instead to reveal these sins of unwarranted rejection and unwarranted acceptance, and in turn, provide a call for representative design, which provides a much stronger foundation for conducting education research that will generalize.

Why would using the term ecological validity to critique research cause any problems? First, the term has never been precisely defined, but has been used in a rather general and vague way. Perhaps the most often-cited advocate of ecological validity has been Ulric Neisser, who in 1976 claimed that “the concept of ecological validity has become familiar to psychologists. It reminds them that the artificial situation created for an experiment may differ from the everyday world in crucial ways. When this is so, the results may be irrelevant to the phenomena that one would really like to explain” (Neisser, 1976, p. 33). Although potentially true, Neisser did not offer specific guidelines on how to achieve generalizability aside from appealing to intuition. Beyond Neisser, Hammond (1998) outlines the vague, multifaceted, and incorrect use of ecological validity by many leaders in the field. Like the above excerpt from Neisser (1976), in other discussions of the concept, some notion of generalizability to one’s target environment is implied, and at times even a specific statement is made about what counts as an ecologically valid experiment. For instance, Bronfenbrenner’s (1977) definition is that “an investigation is ecologically valid if it is carried out in a naturalistic setting and involves objects and activities in everyday life” (p. 515). Generalization may be what we desire to obtain, but these intuitive definitions may be misleading, and even worse, carrying out research “in a naturalistic setting that involves activities in everyday life” may altogether undermine generalizability. A naturalistic setting guarantees nothing, especially given that “naturalistic” is never unpacked—what does it mean?

We are not arguing that education researchers should not be concerned about the degree to which research will generalize in a manner that will improve education practices and student scholarship. However, the sin committed here is that the term ecological validity is used in such a vague manner that it does not provide any guidelines for how to conduct research that is generalizable (for further discussion, see Schmuckler, 2001). Moreover, as an associate editor and as a reviewer for educationally-oriented journals, the senior author of this chapter has too often seen articles blocked by reviewers who boldly state that the research under review is not ecologically valid without providing any explanation about how it does not. We all just assume everyone else knows what ecological validity means, so why fret about explaining how a particular investigation does not meet the mark? Our concern is that otherwise quality articles are sometimes rejected on the basis of such non-constructive critiques. Nevertheless, we also sympathize with those who have erred in this way, because how can we offer a precise statement about the lack of ecological validity, when the concept itself has not been precisely defined?

Second, using the term ecological validity may give the impression that conducting

research in the target ecology ensures that we have met our objective of generalizability. So, if one is interested in improving student metacognition and learning in the classroom, then conducting research in the classroom is by fiat ecologically valid. That this impression is false and can lead to the sin of unwarranted acceptance can be demonstrated using a thought experiment. Imagine that you are conducting an experiment that investigates students' skill at predicting their performance on classroom tests (for a general review, see Hacker, Bol, & Keener, 2008). Such a skill is metacognitive in nature and is apparently useful in helping students to efficiently regulate their learning, because if they accurately predict that they will not do well on a test, they can devote more time for study. You choose to conduct this experiment in your class on Educational Psychology, and realizing the potential importance of accurate metacognition, you encourage all of your students to make accurate predictions. Finally, you examine the absolute accuracy of the students on all four of the classroom exams (for detailed discussion of measures of monitoring accuracy, see Schraw, this volume). The students are at first overconfident, but their predictive accuracy also improves across exams. Experiments like this hypothetical one have yielded some intriguing outcomes about students' metacognitive abilities in the classroom (Nietfeld, Cao, & Osborne, 2006), and our goal here is not to single out any one line of research for criticism. Our main point here is that although this hypothetical experiment was conducted in a classroom, many factors may limit its generalizability to other classrooms. For instance, in contrast to this study, improvements in predictive accuracy may not arise when (a) other teachers do not emphasize the importance of making accurate predictions, (b) a different content area is used in which the students are less interested and motivated to learn, (c) different tests are used (e.g., multiple choice versus essay) that result in different levels of initial absolute accuracy, and so forth.

Our example was meant to make it obvious that conducting research within a naturalistic setting (in this case, a single classroom) does not guarantee generalizability. The more important point, however, is that almost any investigation—regardless of how or where it is conducted—will likely be limited in its generalizability. A similar point was made by Schmuckler (2001), who notes that “concerns with ecological validity are evident in the multiple dimensions of experimental work, including the nature of the experimental setting, the stimuli under investigation, and the observer’s response employed as the measure . . . One consequence of this problem is that concerns with ecological validity can be raised in most experimental situations” (p. 419). Acknowledging this fact will help us make new headway into developing programs of research that will consistently achieve generalizability.

The Challenge for Obtaining Generalizability in Education Science

Ironically, although our thought experiment above was meant to be provocative, it actually falls well short of illustrating the number of factors that can vary in educational settings and hence may influence conclusions and findings from metacognitive research. To emphasize this point, we have generated a list of factors that could in principle influence the outcomes of educational research. These factors are presented in Table 22.1 and refer to different aspects of the environments in which students are taught, learn, and are tested. Any factor that would moderate the effects of an independent variable in an experiment, but was not manipulated within that experiment, by definition limits the generalizability of conclusions from it. We encourage you to generate other factors as well, because we suspect our list is woefully incomplete. Consider just two of these factors and their potential influence on student metacognition and motivation.

Table 22.1 Environmental Factors that may Limit Generalizability

Teachers
Metacognitive orientation
Task and goals
Discourse style with students
Expertise
Background knowledge
Intelligence
Learning environments
Classroom
Resources (computers, internet access)
School locale
Inner city
Rural
College or university
Learning outside the classroom
Home environment
Dormitory
Course materials and content
Examinations
Multiple choice versus essay
Feedback versus no feedback
Comprehensive (cumulative) versus unit-by-unit examinations
Textbook
Narrative versus expository
Text difficulty
Presence/absence of adjunct questions
Sciences
Physical science
Social sciences
Humanities
Art and music

Note: A non-exhaustive list of factors that could influence research outcomes and limit their generalizability.

Textbook Materials and Metacomprehension

Students who can accurately monitor how well they have learned and understood text materials are more effective at learning those materials (for reviews, see Dunlosky & Lipko, 2007; Thiede, Griffin, Wiley, & Redford, this volume). Accordingly, many researchers are attempting to understand metacomprehension accuracy (i.e., the degree to which students' judgments of their text learning predict actual test performance), so as to help students better evaluate their text learning. Perhaps not surprisingly, levels of metacomprehension accuracy are influenced by many factors. For instance, Weaver and Bryant (1995) investigated the impact of the type of text on readers' metacomprehension accuracy by giving them either narrative or expository texts to read. The narrative texts were short stories, such as fairy tales, which were designed to entertain more than to inform. The expository texts were designed to communicate information, like short encyclopedia articles. After reading each text, participants were asked to make metacomprehension judgments for each one (for details about these judgments, see Thiede et al., this volume). Afterwards, they received a test over each text that was composed of multiple-choice questions; half of these questions assessed detailed information, whereas the other half assessed thematic information. The relative accuracy of the judgments was operationalized by an intra-individual correlation between each student's judgment and test performance across the

texts. Accuracy was also computed separately for questions that tapped details and for those that tapped thematic information. Weaver and Bryant found an interaction between the type of text (narrative vs. expository) and type of test question (thematic vs. detailed) used in their study. Specifically, relative accuracy was highest for thematic questions when students read and judged narrative texts, but it was highest for detailed questions when they read expository texts.

Weaver and Bryant also noted that the texts used in their study had not been equated for reading difficulty. Therefore, in two follow-up experiments, they used the Flesch–Kincaid readability index to rate texts on reading difficulty and found that text difficulty was also influencing metamemory accuracy. Specifically, metacomprehension accuracy was highest when texts were at an intermediate difficulty level (approximately a twelfth-grade reading level) in comparison to an easy level (below grade 8) or difficult level (around grade 16).

These effects and interactions have had an impact on metacomprehension theory, but more important for now, they showcase that research conducted in this area that does not consider these influential factors may yield conclusions that may be misleading and limited in generalizability. For example, we may discover an intervention that will boost metacomprehension accuracy, but if our study included texts of only one difficulty level, the degree to which this intervention will successfully generalize to student metacomprehension may be rather limited.

Teacher Discourse Style and Student Motivation

As illustrated in Table 22.1, factors such as text type, question type, and reading difficulty are just a few of the environmental factors that may interact with the main outcomes of education research. As another example, consider teacher discourse style, which inevitably varies between teachers and hence across different classrooms.

Turner, Meyer, Midgley, and Patrick (2003) examined the goal structures established by teachers in sixth-grade classrooms to evaluate their impact on students' positive coping, self-regulation, negative affect following failure, and self-handicapping. Two classrooms were chosen because they were both perceived by students to have the same goal structure (high mastery goals combined with high performance goals), yet the students in the two classrooms exhibited some important differences in affect and behavior. In addition to having the same goal structures, the two classrooms were similar in other ways, too, including demographics, achievement, teacher training, teacher experience, curriculum, and proportion of classroom time spent on various types of learning tasks. Nevertheless, students in the two classrooms differed significantly in their reports of negative affect following failure and in their use of avoidant behaviors such as self-handicapping.

To understand these differences, teacher discourse was analyzed from lessons that had been observed while the same two mathematics lessons were taught in the two classrooms. This analysis revealed differences in the teachers' discourse styles, which seemed to parallel the differences seen in student affect and avoidance behaviors. Whereas both teachers provided instruction that emphasized understanding, one of the teachers did not provide as much support for developing student autonomy. Moreover, whereas both teachers used discourse to try to encourage students, one of the teachers was less consistent in her motivational support, using a higher proportion of negative comments. Put differently, one of the classrooms had teacher–student interactions that were less supportive of student autonomy and more negative regarding motivation, and the students in this classroom self-reported significantly greater levels of self-handicapping and negative affect following

failure. These results point to the importance of considering contextual factors, such as discourse style, when considering the relationship between classroom goal structures and student outcomes. An obvious point here is that metacognitive interventions may have a differential impact in classrooms that are taught by teachers who have a negatively-biased discourse style as compared to those taught by teachers who more consistently adopt a positive style.

To summarize, we would like to emphasize two issues. First, both studies highlight how student outcomes can be moderated by factors that could vary across classrooms. Relevant studies that do not consider these particular factors risk limited generalizability. Second, regardless of their strengths, in terms of generalizability both studies have a similar weakness, because neither examined the full array of factors that could potentially influence student outcomes. Even so, we suspect that according to how some scientists have used the term ecological validity, the first experiment, which was conducted in a laboratory (Weaver & Bryant, 1995), would be chided as not being ecologically valid, whereas the second experiment, which was conducted in classrooms (Turner et al., 2003), would be heralded as being ecologically valid.* It should be equally clear now that neither assessment is appropriate. Moreover, text materials and discourse style are just two of many factors that will vary across classrooms and that could limit the generalizability of education research. This obstacle may not be easily overcome, but to make progress toward this end, we encourage education researchers to consider representative design.

A Call for Representative Design

As we argued above, one reason that using the term ecological validity limits progress in the field is that it has not been precisely defined. Our point was that one popular use of the term ecological validity—when it refers to generalizability—has not been adequately developed. Ironically, when originally introduced, ecological validity had a precise meaning. In fact, Egon Brunswik (1956) coined the term well before it turned into a flaccid moniker for generalizability. For Brunswik, and for many other scientists since, ecological validity refers to the degree to which a distal (or environmental) cue is related to a relevant psychological response. So, consider again the example discussed above concerning students' ability to accurately judge their learning of texts. One distal cue may be the length of sentences in a particular paragraph. If sentence length demonstrates a relatively low correlation with actual learning of the text, this distal cue has low ecological validity. However, if the distal cue of sentence length correlates highly with actual learning, it would have relatively high ecological validity. Ecological validity as defined by Brunswik should be held in particular esteem by those interested in metacognition, because it draws attention to the fact that people will often use a variety of cues to evaluate themselves, such as how well they understand a topic, their ability to understand difficult course content, or the likelihood that they can achieve their learning goals. For each of these dimensions, students may rely on different cues to evaluate themselves (called cue utilization), and the degree to which the cues correlate with the actual dimension being evaluated (ecological validity) will in turn determine the accuracy of students' inferences (for examples of ecological validity in metacognitive research, see Gigerenzer, Hoffrage, & Kleinbölting, 1991; Koriat, 1993). Ecological validity is a powerful concept for understanding the

* Given that this research uses a quasi-experimental design, one might have other reasons to be concerned about the generalizability of results. Our main point here, however, is that conducting research in a classroom does not give it special status as ecologically valid, even if all other aspects of the research are sound.

accuracy of metacognitive assessments; importantly, however, it has nothing to do with generalizability of results to target ecologies.

In contrast, Brunswik's (1956) concept of representative design deals specifically with generalizing to target ecologies, and it does so in a way to clarify the challenges of obtaining generalizability. Representative design refers to a design in which a particular investigation represents the ecology (or environment) outside of the investigation—whether the investigation is conducted in a classroom or in a laboratory—to which we seek to generalize. We flesh out this concept in more detail in the next section. Our main point for now is that over the past three decades the well-defined concept of ecological validity has been confused with representative design. Such confusion in the use of scientific terminology will in part undermine cumulative progress in education science, partly because the incorrect use of ecological validity does not offer definitive guidelines on how to conduct representative research that is most likely to generalize in the intended manner. As more eloquently, and strongly, put by Hammond (1998):

[S]cientists should not take a scientific term that has an established definition and meaning and *use* for over a half century and employ it in an arbitrary fashion that robs it of any meaning whatever. No one has a right to do that. It is bad science, bad scholarship, and bad manners . . . No one should think that this is merely an attempt to resurrect the purity of a term used by an author 50 years ago; it is a term *currently* in frequent (proper) use by many psychologists . . . Therefore its arbitrary change of meaning is a barrier and an affront to those attempting to develop a cumulative science. (p. 2)

We urge all education scientists—and particularly those conducting metacognitive research—to reserve the term ecological validity for applications involving Brunswik's definition: the degree to which a distal (or environmental) cue is related to a relevant psychological response. Besides just using the term as it was originally meant, doing so should ensure that we do not continue to conflate ecological validity with Brunswik's concept of representative design. Although the concept of representative design is itself relatively simple, developing and conducting a representative design in education research may be an ideal that is often difficult to achieve. The stark reality of attaining such an ideal will be more evident as we expand upon representative design.

Beyond Representative Design: The Stark Realities of Conducting Research that Will Generalize

What does it mean to have a representative design in education research? To obtain generalizability across environments, education researchers should begin by describing the environment to which they want their outcomes and conclusions to generalize. In some cases, the environment might be relatively homogeneous, such as if you want your results to generalize to students using a tutor for a particular class. If so, conducting research within that class with that tutor may be sufficient. The idea here is simply that if one is conducting research within the specific environment with which one wishes to generalize, sampling the environment is not an issue and hence generalizability is unnecessary.

In most cases, however, we suspect that researchers desire to generalize to relatively heterogeneous environments, such as to all middle-school classrooms in which genetics is being taught or to all grade-school students learning mathematics. Across classrooms, the environments will vary on many factors (for a subset, see Table 22.1), and to obtain representative design one would have to sample randomly from classrooms in which these

factors vary (and co-vary) in a representative manner. That is, for Brunswik's representative design, we must sample environments just like we were trained to randomly sample participants. For instance, if you are conducting research with fifth-graders in Chicago, you would need to randomly sample from all the fifth-grade classrooms, which differ with regard to teacher abilities, socio-economic make-up of the students, quality of the classroom facilities, and so on. At minimum, perhaps you could develop a description of the degree to which these factors both vary and are intercorrelated across schools in Chicago, and then sample in a manner so that the distribution (and intercorrelations) of these factors are represented in your design. Without doing so, you would run the risk of conducting research that will not widely generalize. This method of obtaining a representative design is akin to stratified random sampling. In the latter case, however, researchers use subject demographics to randomly sample appropriate distributions of subjects within various strata, whereas for representative design, researchers would be concerned about randomly sampling environments.

Why is such sampling necessary to obtain an accurate estimate of the generalizability of our conclusions and findings? To answer this question, consider again the results from Turner et al. (2003), who described how teachers' discourse style influenced student motivation and affect. Assume that in the population of teachers in your target ecology, 80% have a positive discourse style, 10% have a negative style, and the remainder are somewhere in between. You might attempt a metacognitive intervention to boost student comprehension in a classroom, and in doing so, the teacher slated to that class may fall within the 10% of teachers who have a negative discourse style. After you complete the intervention, you report that it had a rather small (but significant) effect on student comprehension—nothing worthy of pursuing any further. If you had randomly sampled 20 classrooms to evaluate your intervention, you would have been more likely to obtain a distribution of teachers that represented the target ecology, and in this case, your effect size—which now capitalizes on the majority of positive teachers—may be substantially larger. This estimate of effect size would better represent the efficacy of your intervention for the target ecology; this effect size would generalize.

Another glance at Table 22.1 may help to further illustrate the difficulties of developing a representative design in education science. In particular, our example above focuses on only one variable—teacher discourse style. However, any of the factors in Table 22.1 may be influential, and as important, these factors do not all vary independently in the target ecology. Thus, to obtain a representative design, our sample of environments must represent not only the distribution of each critical factor within the target population, but it also must represent the inter-correlations among the factors.

Based on this analysis of representative design and our earlier arguments, it should be evident that rarely will any one study meet the mark. Even worse, a series of studies that independently considers various *levels* of a factor may not yield the generalizability that we desire (Maher, 1978). For instance, you may conduct a study to evaluate the efficacy of your metacognitive intervention, and being savvy about teacher influences, you decide to conduct one intervention with a teacher with positive discourse style, and another intervention with a teacher with negative discourse style. As you expected, the intervention yielded a larger effect size (Cohen's $d = 0.70$) with the former teacher, and a much smaller one with the negative teacher ($d = 0.10$). Certainly, the interaction is informative in suggesting the intervention may be less effective with teachers with negative discourse styles. However, the investigation is largely mute with respect to generalizability to the target ecology, at least if you do not know (a) what the distribution of teacher discourse styles is in the population to which you want to generalize and (b) how discourse style interacts (and co-varies) with other unsampled environmental factors.

By now, it should be obvious why Brunswik's (1956) representative design has not been adopted within all domains of psychology. For education researchers who want to generalize to classroom learning, sampling from a representative population of classrooms would be a daunting task. Add to this obstacle the fact that generalization will also require investigating individual differences, the task becomes even more challenging. How can we begin to meet them?

Recommendations

Given the points raised above, one may wonder whether it is feasible to conduct research that is entirely representative of a target ecology. We concede that this ideal may often not be achievable by a single investigator, but even if the ideal cannot be met, we all have much to gain by discussing the implications of representative design for how we conduct and review research. To this end, we offer a small set of recommendations (for other recommendations, see Dhami, Hertwig, & Hoffrage, 2004), which are meant to stimulate conversation and debate about how we can meet the challenges posed by the need for representative design to achieve generalizability.

First, we should stop using ecological validity to mean generalizability, and instead use the term to mean ecological validity as developed by Brunswik (1956). In contrast to Hammond (1998), we are not concerned about "bad manners," because for researchers interested in metacognition, the distinction goes well beyond semantics and poor taste. For metacognition research, Brunswik's concept of ecological validity is essential for understanding monitoring and control processes that can influence student scholarship. Students use cues in their environments (both internally and externally) to evaluate their progress and to make decisions about how to regulate their learning, and hence the ecological validity of these cues will influence their success. The concept of ecological validity can guide research on metacognition, but it is much less likely to do so when we use it to mean something else.

Second, we should be explicit about the ecology that our research is targeting. Perhaps this is obvious in most articles, because after all, an article about self-regulated learning in a college biology class should at least be targeting other biology classes at the same level. As reviewers of psychological literature, we should be more sympathetic to these issues and, at a minimum, should always explain ourselves when we state, "this research is not representative." When you are tempted to critique a paper in this manner and hence run the risk of committing the sin of unwarranted rejection, ask yourself, "How does this research fail to meet a standard of generalization expected by the field?" For that matter, one excellent question for the entire field would be, "Exactly what is the standard of generalization we should be expecting?" We cannot answer this question confidently without a larger discussion within the field, but expecting all research to be built on representative design is setting the standard much too high.

Third, a goal of education research should be to continue to understand the environments in which student learning and performance occurs. The list presented in Table 22.1 is not exhaustive, yet it already comprises a sizeable number of factors that could plausibly influence student learning. Which of them should we choose to investigate? Which ones will be influential, and importantly, which ones are not? Answering these questions is vital, because doing so will limit the number of possible relevant factors to sample from and hence make the reality of conducting representative design more feasible. Systematic collection of classroom demographics, which could support the development of a theory of these environments, could guide decision making about which factors to sample when designing education research.

Fourth, education researchers have developed many programs that successfully improve student learning. These successes originated from research that was not representative, but instead from individual researchers, who were investigating a promising—but untested—method usually in a single classroom. Representative designs will require sizeable funds to complete, and perhaps rightly so, substantial grants are not available for risky interventions. So, perhaps ironically, representative designs in education research will often be fostered by progressive research that is non-representative with respect to the target ecology; in fact, such systematic research will likely be the root of all representative designs in education research. We should continue conducting our research—in the laboratory and in the classroom—but in doing so, attempt to make our designs as representative as possible and to understand their limitations.

Conclusion

Even if all sins cannot be forgiven, the sins committed in the name of ecological validity can easily be forgotten as we move toward a research culture that seeks to achieve generalizability through the use of representative design. By shifting our focus to representative design, metacognitive research aimed at improving student scholarship will reap at least two benefits. First, ecological validity is the degree to which an environmental cue predicts a proximal psychological response. Defined in this manner, ecological validity becomes a useful tool for understanding the biases in students' judgments and why they are often inefficient at regulating their learning. Second, and as important, Egon Brunswik (1956) precisely defined representative design, and adopting such a design will ensure research outcomes generalize to target ecologies. Discussing how to achieve this design ideal is bound to promote genuine advances that will leave no child behind, regardless of his or her personal characteristics or learning environments.

Author Notes

Thanks go to Katherine Rawson for discussion and feedback on this chapter. Development of this chapter was partially supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305H050038 to KSU. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

- Araújo, D., Davids, K., & Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comments on Rogers, Kadar, and Costall (2005). *Ecological Psychology, 19*, 69–78.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32*, 513–531.
- Brunswik, E. (1956). *Perception and representative design of psychological experiments*. Berkeley: University of California Press.
- Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin, 130*, 959–988.
- Dunlosky, J. & Lipko, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*, 228–232.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.
- Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration.

- In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 429–455). New York: Taylor & Francis.
- Hammond, K. R. (1998). *Ecological validity: Then and now*. Retrieved from <http://www.brunswik.org/notes/essay2.html>
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*, 609–639.
- Maher, B. A. (1978). Stimulus sampling in clinical research: Representative design reviewed. *Journal of Consulting and Clinical Psychology*, *46*, 643–647.
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. New York: W. H. Freeman and Company.
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning*, *1*, 159–179.
- Petrinovich, L. (1989). Representative design and the quality of generalization. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 11–24). New York: Cambridge University Press.
- Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, *2*, 419–436.
- Turner, J. C., Meyer, D. K., Midgley, C., & Patrick, H. (2003). Teachers' discourse and sixth graders' reported affect and achievement behaviors in two high mastery/ high performance mathematics classrooms. *The Elementary School Journal*, *103*, 357–382.
- Weaver, C. A., III & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition*, *23*, 12–22.

Index

- academic ability and knowledge monitoring
114–15
- adaptive support 384, 404
- Advanced Distributed Learning 121
- affect 368–9, 375–6, 407
- agency 1–2, 260, 261
- Ainley, M. et al. 302
- Aleven, V. et al. 111, 272, 387–9, 398
- Alexander, J. M. et al. 210
- Allwood, C. et al. 224*t*, 230, 420
- American Psychological Association (APA) 227
- Anders, P. 247
- Anderson, J. R. 390
- Anderson, M. C. M. 89*t*, 92*t*, 98, 288, 422
- Annacone P. 306
- Araújo, D. et al. 430
- Arenas, A. et al. 265
- Arroyo, I. et al. 397
- Asimov, I. 18
- AutoTutor 367, 371–2, 376
- Azevedo, R. et al. 122, 384
- Baker, J. M. 89*t*
- Baker, L. 7, 241, 242
- Baker, R. et al. 393, 403
- Bandura, A. 120–1, 199, 303–4
- Barber, B. M. 236
- Baron-Cohen, S. 29
- Barresi, J. 2
- Baumann, J. F. 242
- Beall, L. 241
- Beck, I. L. et al. 8, 13, 38
- belief systems 263–4
- Bell, T. 116
- Bem Sex Role Inventory (BSRI) 230–1
- Benjamin, A. S. et al. 283
- Bereiter, C. 50, 131, 137–9, 158, 169, 199
- Berliner, D. C. 244, 245
- Berlyne, D. E. 49, 50
- Best, R. et al. 65
- Beyer, S. 230, 234
- Billings, J. 236
- Biswas, G. et al. 342
- Bjork, R. A. 284, 285, 286, 294–5, 390, 392
- Bol, L. et al. 422
- Borkowski, J. G. 211
- Bracewell, R. J. 157
- Bransford, J. et al. 241, 244, 249, 250, 383
- Brekke, N. 279
- Brenna, B. A. 30, 31
- Brewer, W. F. et al. 423
- Britner, S. 222
- Bronfenbrenner, U. 431
- Brophy, J. 119
- Brown, A. 176, 383
- Brown, A. L. et al. 7, 9–10, 29, 34, 107
- Brown, R. et al. 35
- Bruer, J. 157, 390
- Brunswik, E. 435, 436, 438, 439
- Bryant, D. S. 94*t*, 101, 433–4, 435
- Bryk, A. S. 250
- BSRI (Bem Sex Role Inventory) 230–1
- Buchmann, M. 241
- Buettner, G. 311
- Burson, K. A. et al. 423
- Burton, S. 221
- Butler, A. C. et al. 287
- Caldeira, M. H. et al. 55
- Carter, K. 241, 244
- Carver, C. S. 158
- Carver, R. P. 12
- CATS (Close Analysis of Texts with Structure)
37
- Chall, J. S. 38
- Cheng, R. 393
- Chi, M. T. H. et al. 365, 373, 384, 385, 389
- Children's Sex Role Inventory (CSRI) 227–8
- Church, M. 116, 117
- Clark, R. E. 112
- Clay, M. M. 38
- Clifford, M. M. 369
- Close Analysis of Texts with Structure (CATS)
37
- Cochran-Smith, M. 242
- Cognitive Strategy Instruction for Writing
(CSIW) Program 142
- Collaborative Strategic Reading (CSR) 35
- Collins, A. 179
- computer tutors 367, 371–2, 376

- Conati, C. et al. 385–6, 389–90
 concept mapping and cue-access 99–100
 conditional knowledge 133, 135
 confidence—gender and cross-cultural
 differences 221–36, 422; characteristics of
 populations 230–1; confidence
 discrimination 228–9, 229*f*, 266; cultural
 influences on gender and confidence 233;
 culture 227, 231, 232*f*; implications for
 teachers 233–5, 234*t*; judgments of
 confidence 221; mathematics 221–3, 229,
 233; measuring confidence 222–8:
 (calibration 223; confidence discrimination
 226; culture: country or ethnicity 227;
 experimental or classroom setting 226–7;
 gender: sex or instrument 227–8; global
 confidence 222; item-specific confidence
 222–6, 224–5*t*; significance tests or CAQs
 226); research context 226–7, 232–3; self-
 concept and self-efficacy 221–2; subject area
 229–30
 Conway, M. A. 2
 Corbett, A. T. et al. 385, 389, 390
 Corno, L. 247
 Cornoldi, C. et al. 208
 Correnti, R. 243
 Costa, J. et al. 48
 Cox, B. E. 31
 Craig, S. D. et al. 368
 Csikszentmihalyi, M. 369
 CSIW (Cognitive Strategy Instruction for
 Writing Program) 142
 CSR (Collaborative Strategic Reading) 35
 CSRI (Children's Sex Role Inventory) 227–8
 cue-access: and concept mapping 99–100; and
 self-explanation 98–9
 cultural differences: children's writing 140;
 knowledge monitoring assessment (KMA)
 research 117–18; mathematics 221–2, 233;
 measuring metacognitive judgments 422; *see*
 also confidence—gender and cross-cultural
 differences
 Davidson, G. R. 211
 Davis, S. L. 120
 Day, C. 248
 De Clercq, A. et al. 212
 Deci, E. 301
 declarative knowledge 133, 134–5
 delayed-generation effect 97–8
 Dennett, D. 344
 Dennett, D. C. 2
 Dennison, R. S. 116
 Desoete, A. 122, 213
 Dewey, J. 182
 Dhami, M. K. et al. 438
 dialogical self 2
 Didion, J. 161
 Dignath, C. et al. 310, 311
 Dillon, A. 320
 Dillon, J. T. 48
 Direct Explanation (DE) approach 38–9
 directed retrospection technique 164
 D'Mello, S. K. et al. 368
 Doherty, M. 40
 Dole, J. A. 12
 Dougherty, M. R. P. 421, 422, 423
 Dowker, A. 206
 Duffy, G. G. et al. 10–11, 27, 37, 38, 241, 245,
 246, 247, 248
 Duffy-Hester, A. M. 242
 Dunlosky, J. et al. 86, 88, 89*t*, 92*t*, 95, 285,
 286–7, 289, 290, 291
 Dunn, J. C. et al. 04 426
 Dweck, C. 302
 ease of learning judgments (EOLs) 416; *see also*
 measuring metacognitive judgments
 Eccles, J. S. 301
 ecological validity 430–9; environmental
 factors 432–5, 433*t*; generalizability 431,
 432–5, 436–8; representative design 435–6,
 438, 439; teacher discourse style and student
 motivation 434–5; textbook materials and
 metacomprehension 433–4;
 recommendations 438–9
 eHElp 269
 Ekman, P. 368, 375–6
 Elliott, A. 116, 117
 Emig, J. 132
 Englert, C. S. et al. 142
 EOLs *see* ease of learning judgments
 EPA2000 212
 Epstein, W. 86, 420
 Ericsson, K. A. et al. 264
 error self-correction 265, 270, 390–3, 391*f*,
 392*f*
 Everson, H. 109, 112, 113, 209, 212
 executive control 7, 10
 executive function 134
 eye-tracking technology 164–5
 Facklam, M. 19
 false belief tasks 28, 40
 Farrar, M. J. et al. 40
 feeling of knowing judgments (FOKs) 282, 416;
 with hypermedia 325, 326*t*, 327, 331–2, 334;
 see also measuring metacognitive judgments
 Feltz, D. 236
 Ferguson, W. 179
 Fielding, L. 12
 Finn, B. 287, 288, 294
 Finney, S. J. 120
 Fischhoff, B. 283–4
 Fishman, B. J. et al. 250–1
 Fitzgerald, J. 166
 Flammer, A. 49
 Flavell, J. H. 27, 107, 133, 278, 415
 Fletcher, J. D. 121
 Flores, M. 248

- Flower, L. 132, 133, 136–7, 140, 157, 163, 164, 169
 FOKs *see* feeling of knowing judgments
 Francis, B. 235, 236
 Franco-Watkins, A. M. 423
 Frederiksen, J. R. et al. 180, 187, 192, 199
 Freebody, P. R. 211
 Friesen, W. V. 368, 375–6
 Frommer, F. D. 49
- Gabbard, R. 320
 Galda, L. 241
 Gallagher, M. C. 246
 Garrison, J. 245
 Gavelek, J. R. 22, 47, 50
 Geary, D. C. 206, 208
 Gelman, R. 341
 gender *see* confidence—gender and cross-cultural differences
 generalizability 431, 432–5, 436–8
 Generalized Estimating Equation (GEE)
 analyses 64
 Geometry Cognitive Tutor 387–9, 388*f*, 397, 400, 401
 Gerrity, H. 110, 112–13
 Gersten, R. et al. 12
 Gillstrom, A. 422
 Ginsburg, H. P. 206, 209
 Glaubman, R. et al. 32–3
 Glenberg, A. M. et al. 86, 282, 284, 420
 Goleman, D. 375
 Graesser, A. C. et al. 47, 48, 49, 50, 51, 53, 54, 55, 364, 365, 367, 370, 375
 Graham, S. 35, 140, 141, 142, 143, 144, 145, 149, 235
 Grant, H. 302
 Griet, T. 116, 123
 Griffin, T. D. et al. 90*t*, 99, 100–1, 288
 Grossman, H. 235
 Grossman, P. L. et al. 247
 Grossman, S. 235
 gStudy 268–9
 Gu, H. 422, 426
 Guice, S. 241
- Hacker, D. J. 22, 133
 Hadwin, A. et al. 261–2, 261*f*, 263, 292, 322, 323, 324, 330
 Hammerton, L. 271–2
 Hammond, K. R. 431, 436, 438
 Hanlon, R. 119
 Harackiewicz, J. M. et al. 302
 Harris, K. R. et al. 35, 140, 141, 142, 143, 145, 146, 149
 Harris, R. 154–5
 Harter, S. et al. 227
 Hartman, H. 213, 214
 Hattie, J. 300
 Hausmann, R. G. M. 389
 Hawley, W. 247
- Hayes, J. 132, 133, 136–7, 140, 154, 157, 163, 164, 169
 He, Y. 248
 Heffernan, N. T. 397
 help-seeking 111–12, 265, 271–2, 303, 397–402, 399*f*, 400*f*
 Help Tutor 272
 Hidi, S. 301–2
 Himley, M. 157
 Hong, E. et al. 120
 Huberman, M. 248
 Hunt, J. E. 422
 hypermedia *see* self-regulated learning with hypermedia
- I-Help 271
 imagery 302–3
 Impara, J. C. 108
 IMPROVE 213–14
 information-processing theory of SRL 320–3
 Informed Strategies for Learning (ISL) 10, 33–4
 Institute of Education Sciences (IES) 279
 intelligent tutoring systems 3, 383–408;
 adaptive support 384, 404; analysis of interventions 404–7, 406*t*; error self-correction 390–3, 391*f*, 392*f*; gaming and affect, motivations and attitudes 402–3; Geometry Cognitive Tutor 387–9, 388*f*, 397, 400, 401; goals for robust learning 384–5, 385*f*; help seeking 397–402, 399*f*, 400*f*; Progress Tips 397; reducing gaming the system 393–7; role of affect in metacognitive choices 407; Scooter the Tutor 393–7, 394*f*, 396*f*; self-explanation 385–90, 388*f*; shallow reasoning 386–7, 386*f*; static support 384, 404; tutoring and assessing metacognition 407–8
 interactive metacognition 340–1; *see also* intelligent tutoring systems; Teachable Agents; tutoring: meta-knowledge
 Ishiwa, K. et al. 52, 54
 ISL *see* Informed Strategies for Learning
 iSTART (Interactive Strategy Training for Active Reading and Thinking) 67, 68*f*, 73
- Jackson, J. W. 273
 Jacoby, L. L. 283
 James, W. 1, 164
 Jee, B. et al. 90*t*
 Jobst, J. 320
 Johnston, P. 12
 Jones, L. G. 223, 225*t*, 229, 235
 Jones, L. P. 223, 225*t*, 229, 235
 Jonsson, A. 224*t*, 230
 Juckes, T. J. 2
 judgments of learning (JOLs) 262, 416;
 accuracy 280–1; delayed-JOL effect 88, 285–8; with hypermedia 325, 326*t*, 327, 331–2; and study time 290–1; *see also* measuring metacognitive judgments

- Juliebö, M. et al. 30–1
 Juslin, P. et al. 423, 424–5, 426
- Karabenick, S. A. 48
 Kelemen, W. L. et al. 422
 Kellogg, R. T. 156, 157, 164
 Kendeou, P. et al. 36
 Keren, G. 421
 Kim, B. 227
 Kimball, D. R. 286, 287
 King, J. F. et al. 285
 Kinnunen, R. et al. 31
 Kintsch, W. et al. 12, 55–6, 88, 97
 Kitsantas, A. 305, 306, 307
 Kluwe, R. H. 1, 157
- knowledge: conditional knowledge 133, 135;
 declarative knowledge 133, 134–5;
 metacognitive knowledge 278–9; meta-
 knowledge 176, 178–86; procedural
 knowledge 133, 135
- knowledge monitoring assessment (KMA)
 framework 4, 107–12, 212; accuracy
 110–11, 110*f*; assessment 109–10; help
 seeking 111–12; hierarchy of metacognitive
 processes 109, 109*f*, 122
- knowledge monitoring assessment (KMA)
 research 112–21; cross-cultural studies
 117–18; domain specificity 115; general
 academic ability 114–15; mathematics
 113–14; motivation and metacognition
 116–20; motivational goal orientation
 119–20; reading 113; self-regulation and
 metacognition 120–1; self-reported
 metacognition 115–16; training in
 metacognitive knowledge monitoring 123;
 future directions 121–3
- Koedinger, K. R. 385, 387–9, 390, 398
 Koivula, N. et al. 224*t*, 228, 229, 230, 234
 Koriat, A. et al. 95, 280–1, 283, 284, 285, 423
 Kornell, N. 290, 291, 292, 294–5
 Kramarski, B. 213–14
 Kuhn, T. S. 155, 170
- language and theory of mind 28–9, 40
 Latent Semantic Analysis (LSA) 64–5, 76
 Learning and Study Skills Inventory (LASSI)
 115–16
- learning disabilities and writing 138, 142; *see*
also mathematical learning disabilities
- Leggett, E. 302
 Lepper, M. R. 375
 Levin, B. 248
 Lewis, M. M. 49
 Light, R. J. 119–20
 Lin, L. et al. 86, 90*t*, 101, 423
 Lin, X. D. 241
 Lin, X. et al. 241
 Lindauer, B. K. 134
 Lipko, A. R. 86, 88
 Little, J. W. et al. 241
- Lloyd, J. et al. 222–3, 225*t*, 228
 Lockl, K. 29
 Logo 356
 Lortie, D. L. 247
 LSA *see* Latent Semantic Analysis
 Lucangeli, D. 208
 Luckin, R. 271–2
 Lundeberg, M. A. et al. 223, 224*t*, 225*t*, 226,
 227, 228, 230, 231, 236, 422
 Lytle, S. L. 242
- McCartney, K. 299
 McCormick, C. B. 134, 136
 McCutchen, D. 154
 McDaniel, M. A. 93*t*, 99, 102, 267, 289
 McKeown, M. G. et al. 8, 13, 38
 McMahan, C. L. 49, 50, 53
 McNamara, B. E. 206
 McNamara, D. S. et al. 64, 65, 66–7, 69, 73,
 74–6, 75
- Maehr, M. L. 116, 118, 120
 Magliano, J. P. et al. 64, 65–6, 70, 73
 Magnussen, R. 393
 Magnusson, S. J. 186–7
 Maki, R. H. et al. 86, 90–2*t*, 91*t*, 92*t*, 422
 Maloch, B. et al. 246, 247
 Many, J. E. 246
 Markman, E. M. 31
 MARSII (Metacognitive Awareness of Reading
 Strategies Inventory) 73
 Martinez-Pons, M. 121
 Mathan, S. A. 390
- mathematical learning disabilities (MLD) 3,
 206–15; cognitive underpinnings 208–9;
 conceptual model 207–8, 207*f*; deficit
 hypothesis 211; definition 206; impact of
 deficits 209; intelligence 210; metacognitive
 assessments 211–13; metacognitive skills
 209–10; metacognitive underpinnings
 209–10; number knowledge deficits 208–9;
 procedural deficits 208; retardation
 hypothesis 210–11; semantic memory
 deficits 208; STICORDI devices 214–15;
 training of metacognition 213–14; visual
 spatial deficits 208
- mathematics: cultural differences 221–2, 233;
 gender differences 221, 222–3, 229, 233;
 knowledge monitoring 113–14; *see also*
 mathematical learning disabilities
- Mayer, R. E. 99
- measuring metacognitive judgments 4, 415–27;
 absolute accuracy 417–19, 418*t*; concurrent
 judgments 416, 416*t*, 417; cultural
 differences 422; hybrid scores 421;
 individual differences 422; metacognition
 415–16; metacomprehension 415;
 metamemory 415; outcome scores 421,
 424–5; prospective judgments 416, 416*t*;
 relative accuracy 417, 418*t*, 419–21;
 reliability 424; research overview 421–3;

- retrospective judgments 416, 416*t* 417; task parameters 422–3; text and test parameters 423; verbal ability 422; working memory 422; recommendations for future practice 425–6; recommendations for future research 426–7
- Meck, E. 341
- Meece, J. L. et al. 235, 236
- Meister, C. 34
- memory: long-term memory 136, 137; metamemory 415; for past test 288; semantic memory deficits 208; working memory 137, 422
- meta-affect 368–9, 375–6
- metacognition: definitions 27, 278, 383, 415–16; knowledge about cognition 133–4; origins 7; regulation and control of cognitive activity 134; and theory of mind 28–9
- Metacognitive Awareness of Reading Strategies Inventory (MARS-I) 73
- metacognitive control 290–5; definition 290; and monitoring accuracy 292–5; strategy selection 292; study time allocation 290–2
- metacognitive knowledge 278–9
- metacognitive monitoring 2, 85–102, 279–89; accuracy of judgments 280–1, 292–5; demands during reading 100–1; heuristics for improving accuracy 285–9: (debiasing incorrect heuristics 285; knowledge of test 95, 289; memory for past test 288; retrieval attempt 285–7; summarization 288–9); heuristics producing inaccuracies 280–5: (association 284; current knowledge 283–4; familiarity 281–2; fluency 282–3; illusions of knowing 279, 284–5); improving monitoring accuracy 86; level of understanding 96–100: (concept mapping and cue-access 99–100; delayed-generation effect 97; self-explanation and cue-access 98–9); metacomprehension accuracy: review 88, 89–94*t*; monitoring judgments 279–80; poor monitoring accuracy when learning from text 86–8
- metacommunication 366–8, 374–5
- metacomprehension 267–8, 415
- Metacomprehension Strategy Index (MSI) 71–3, 71*t*, 72*t*
- meta-knowledge 176, 178–86
- metamemory 415
- Metcalfe, J. et al. 281, 282, 286, 287, 288, 290–1, 291–2, 294
- Mevarech, Z. R. 213–14
- Meyer, D. K. 368–9
- Meyer, H. A. 116
- Miller, G. E. 11
- Millis, K. K. 64, 65–6
- Misfeldt, M. 393
- Mitrovic, A. 389
- Miyake, N. 48, 54, 365
- MLD *see* mathematical learning disabilities
- Moely, B. E. et al. 38
- Moore, E. C. 156
- Moos, D. C. 122
- Moreno, R. 270
- Morrow, L. M. et al. 245
- Mosenthal, J. et al. 245
- Motivated Strategies Learning Questionnaire (MSLQ) 115–16, 121
- motivation: goal orientation 119–20, 302; and metacognition 116–20; outcome expectancies 301; task analysis 301; task interest/valuing 301–2; technology for self-regulated learning 272–4; *see also* self-regulated learning—metacognition and motivation
- MSI *see* Metacomprehension Strategy Index
- Murray, R. C. 397
- Narens, L. 161
- narrative self 2
- Nathan, R. 120
- National Assessment of Educational Progress 131
- National Commission on Writing 132
- National Reading Panel (NRP) 9, 11, 14, 32, 245
- National Research Council (NRC) 14, 177
- Neisser, U. 431
- Nelson, T. O. et al. 86, 110–11, 161, 286–7
- Nesbit, J. C. 268–9
- Nietfeld, J. L. et al. 110–11, 422, 425, 426
- Njoku, H. U. 111, 118, 120
- No Child Left Behind 243, 246
- Norman, D. A. 48, 54, 365
- Nystrand, M. 132, 154, 157, 170
- Oakhill, J. 50
- Odean, T. 236
- Okita, S. Y. 341
- Olde, B. 55
- Olin, J. T. 92*t*
- Omizo, M. 227
- O’Neil, H. F. et al. 121
- O’Neill, J. K. 112
- Otero, J. 50, 55–6
- other regulation 7, 9
- Ozuru, Y. et al. 65
- Paas, F. G. W. C. 396
- Pajares, F. 222, 227, 230, 308
- Palincsar, A. S. 9–10, 34, 186–7
- Pallier, G. 224–5*t*, 226, 229–30, 235
- Papert, S. A. 356
- Paris, A. H. 248
- Paris, S. G. et al. 10, 33, 134, 242, 248
- Parsons, S. A. 246, 247, 250
- Pashler, H. et al. 259, 274, 279
- Pearson, P. D. 12, 243, 246, 250
- pedagogical agents *see* Teachable Agents (TA)
- Pierce, C. S. 156
- Perin, D. 143, 235

- Perner, J. 40
 Perry, N. E. et al. 250
 Perry, P. J. 72–3, 75, 77
 Person, N. K. et al. 364, 365, 367, 370, 372, 373, 374
 Petrinovich, L. 430
 Phillips, L. D. 233
 Piaget, J. 47, 51, 52, 56
 Pierce, B. H. 92*t*
 Pintrich, P. R. et al. 107, 110, 113, 115, 116, 118, 120, 262, 383, 403
 Plato 156
 Popper, K. 184
 PREG model 50–1
 Pressley, M. et al. 10, 34, 37, 120, 245, 246
 Prior, P. 140, 154
 procedural knowledge 133, 135
 Programme for International Student Assessment (PISA) 221
 Progress Tips 397
 Puncocchar, J. et al. 231
- question generation 47–57; anomaly detection 48; anomaly hypothesis 49; knowledge clash hypothesis 48; knowledge deficit hypothesis 48; mechanisms 47–9; obstacle+goal hypothesis 48–9; obstacles 50–3; reader's knowledge 54–7; (initial representation 55–6; representation goal 56–7); reading goals 53–4; situation model 49
 Questioning the Author (QtA) 13; *see also* reading comprehension comparative study
- R-SAT *see* Reading Strategy Assessment Tool
 Ram, A. 49, 51–2, 54
 RAND Study Reading Group 26
 Randi, J. 247, 248
 Raphael, T. E. 22, 47, 50
 Raudenbush, S. W. 250
 Rawson, K. A. et al. 89*t*, 92*t*, 95, 101
 reading: knowledge monitoring 113; *see also* reading comprehension; reading comprehension comparative study; reading comprehension with primary students; self-explanation in reading
 reading comprehension 2, 7–13; cognitive processing 12–13; Informed Strategies for Learning (ISL) 10; metacognition and reading 7–9; modeling 10–11; Questioning the Author (QtA) 13; Reciprocal Teaching (RT) 9–10; strategies literature 9–11; strategy instruction 11–12; Transactional Strategies Instruction (TSI) 10
 reading comprehension comparative study 13–23; classroom lesson discussion 19–21; measures and outcomes 15–16; metacognition 22–3; students' comprehension 16–19
 reading comprehension with primary students 26–41; assessment methods 30–1; cognition 27; definition 26; empirical studies 29–32; metacognition 27; metacognitive reading instruction 30, 32–8; (instructional programs 33–5; integrating listening and comprehension instruction 35–8; research 32–3); skills 26, 27; strategies 26–7, 30–1; teaching metacognition to teachers 38–9; theory of mind 28–9, 40
 Reading Strategy Assessment Tool (R-SAT) 66, 78
 Reading Strategy Checklist 71, 71*t* 72
 Reciprocal Teaching (RT) 9–10, 34
 Reder, L.M. 281–2
 reflection 177, 185, 191, 191*f* 242
 Region of Proximal Learning (RPL) model 293
 Reisenberg, R. 135, 135*t*, 139–40
 Renkl, A. 112
 Renninger, K. A. 301–2
 representative design 435–6, 438, 439
 Richardson, V. 247
 Risko, V. J. et al. 242
 Ritter, F. E. 281–2
 Rodgers, E. M. 246
 Roedel, T. D. 423
 Roehler, L. R. et al. 10–11, 38, 245, 246
 Rogers, L. 144
 Rohman, G. 132
 Ronnberg, J. 422
 Rosenshine, B. 34
 Rosenthal, R. 236
 Rowan, B. 243
 Royer, J. M. et al. 108
 RPL (Region of Proximal Learning) model 293
 RT *see* Reciprocal Teaching
 Ruan, J. 31–2
 Rubin, D. B. 236
 Ruffman, T. 28–9
 Ryan, R. M. 301
- Sandora, C. et al. 13
 Saunders, J. et al. 233
 Scardamalia, M. 50, 131, 137–9, 158, 169, 199
 Scarr, S. 299
 Schank, R. C. 49, 53
 Scheier, M. F. 158
 Schmidt, R. A. 390
 Schmitt, M. C. 11
 Schmuckler, M. A. 430, 431, 432
 Schneider, W. 29
 Schoenfeld, A. H. 177, 345, 383
 Schofield, J. 393
 Schommer, M. et al. 263
 Schraw, G. 108, 111, 116, 423, 424, 425
 Schunk, D. H. 307
 Schwonke, R. et al. 269
 science education—developing students' meta-knowledge 186–201; guided inquiry and reflective assessment 187–9, 188*f*; learning through inquiry about inquiry 198–9;

- notebook texts 186–7; role playing 194–8, 197*f*; Web of Inquiry 189–94; *see also* ecological validity
- scientific inquiry 3, 175–86; Inquiry Cycle 179; meta-knowledge 176, 178–86: (for data analysis 183–5; meta-investigation knowledge 181–3; meta-questioning knowledge 180–1; meta-theoretic knowledge 179–80; metacognitive control of inquiry processes 185–6); metacognition and teaching 177–8; metacognitive capabilities 176; models 176*n*; planning and monitoring 177; reflection 177, 185; self-regulation 185; theories 176*n*; *see also* science education—developing students’ meta-knowledge
- Scooter the Tutor 393–7, 394*f*, 396*f*
- self-concept 1–2, 4, 221–2
- self-control 302–3
- self-efficacy 135–6, 139, 177–8, 200–1, 221–2, 272–3, 301, 307–8
- self-evaluation 303–4, 308
- self-explanation 385–90, 388*f*
- self-explanation in reading 2, 60–78; and cue-access 98–9; defined 61–3, 63*t*; Generalized Estimating Equation (GEE) analyses 64; levels of comprehension and reading processes 63–6; metacognition and comprehension 70–3; and reading strategies 73–6; think aloud protocols 60, 62, 64–6, 74; training 66–70
- Self-Explanation Reading Training (SERT) 67, 69–70, 69*f*
- self-observation 303
- self-reaction 304
- self-reflection 185
- self-regulated learning (SRL) 3, 259–74; agency 260, 261; belief systems 263–4; computing technologies 265–74; gateways to successful cognition 260, 260*t*; help-seeking 265, 271–2; judgments of learning (JOL) 262; learning from errors 265, 270; metacognition as a production system 259–60; phases of self-regulated learning 261–2, 261*f*, 322; self-efficacy 272–3; study tactics 263, 268–9; “what works” 264–5
- self-regulated learning, metacognition and motivation 299–313; cyclical phase model 300–5, 300*f*: (forethought phase 300–2; performance phase 301, 302–3; self-reflection phase 301, 303–5); intervention study 308–10; microanalytic measures 305–6; proactive self-regulators 306–7; self-efficacy and self-evaluative calibration 307–8; software 272–4; evaluation of training 310–12
- self-regulated learning with hypermedia 3, 319–35; information-processing theory of SRL 320–3; learning with hypermedia 320–1; monitoring guidelines 334–5; monitoring processes 324–34, 326*t*: (deployment of self-regulatory processes 328–34, 329*f*, 330*f*, 331*t*, 332*t*, 333*f*)
- Self-Regulated Strategy Development (SRSD) 35, 142–50; critical characteristics 146–7; development 145–6; effectiveness 143–5: (single-subject design studies 144–5; true- and quasi-experimental design studies 143–4); generalization and maintenance 149; metacognition and strategies instruction 147–9; process of instruction 147, 148*t*; research needs 149–50
- self-regulation 7, 185, 242, 250; in children’s writing 135–6, 135*t*, 139, 140; and metacognition 120–1; *see also* self-regulated learning (SRL); self-regulated learning, metacognition and motivation; self-regulated learning with hypermedia; Self-Regulated Strategy Development (SRSD)
- Serra, M. 91*t*
- SERT *see* Self-Explanation Reading Training
- Shalev, R. et al. 208
- shallow reasoning 386–7, 386*f*
- Shaughnessy, J. J. 226
- Shavelson, R. 246
- Sheeran, P. 119
- Shulman, J. 247
- Shulman, L. 241, 247
- Siegler, R. S. 346
- Sinatra, G. M. et al. 12, 22
- Sitko, B. M. 132
- Skelton, C. 235, 236
- Slade, L. 28–9
- Smith, S. M. 92*t*
- Snow, C. et al. 245
- Son, L. K. 287, 290–1, 422
- Soto, G. 17–18
- Souchay, C. et al. 422
- Spellman, B. A. 286
- Spillane, J. P. et al. 241
- Sprenger, A. 421, 422, 423
- Squire, J. R. 38
- SRL *see* self-regulated learning
- SRSD *see* Self-Regulated Strategy Development
- Stahl, E. et al. 263
- static support 384, 404; for self-explanation 385–90, 388*f*
- Stavrianopoulos, K. 117
- Sternberg, J. 213
- Sternberg, R. J. 114
- Stevens, T. et al. 231
- STICORDI devices 214–15
- Stone, N. J. 308
- task analysis 301
- task environment 136, 137
- Taylor, B. M. et al. 245
- Teachable Agents (TA) 341, 342–56, 342*f*; adopting cognition of the agent 346, 347–8, 347*f*, 348*f*; data mining methodology for analyzing learning choices 352–5, 353*t*, 354*f*;

- monitoring one's own thoughts in an agent 343–6, 345*t*, 346*t*; regulating cognition for the agent 348, 349–52, 349*f*, 350*t*; science content 350–1; transfer of learning 351, 352*t*; Triple-A-Challenge Gameshow 343, 344*f*
- teachers as metacognitive professionals 4, 240–51; dispositional perspective 248; effective teachers 240–1, 244–6; expert teachers 245; impact on student achievement 249; metacognition and teaching 177–8, 241–7; methodological problems 244; nomenclature problems 242, 249; personal practical theories 248; professional development 247–8; scaffolding 246; situational problems 243; teacher decision-making 246; teacher identity 248; vision 248; future research 248–51
- teachers: metacognition instruction for 38–9
- technology for self-regulated learning 265–74; eye-tracking technology 164–5; facilitating metacomprehension 267–8; for help-seeking 271–2; for learner data collection 266; for learning from errors 270; for learning tactics and strategies 268–9; motivation for self-regulation 272–4; software use 266–7; *see also* computer tutors; intelligent tutoring systems; self-regulated learning with hypermedia; Teachable Agents
- Theme Scheme 36
- theory of mind 28–9, 40
- Thiede, K. W. et al. 85, 92–3*t*, 92*t*, 97, 98, 99–100, 102, 189*t*, 287, 288, 289, 290, 291, 293, 422
- think aloud protocols 60, 62, 64–6, 74, 122, 164, 329*f*
- ThinkerTools Inquiry Curriculum 187
- Thomas, A. K. 93*t*, 99, 102, 267, 289
- Thorpe, P. K. 211
- Timperley, H. 300
- Tobias, S. et al. 109, 110, 112–13, 116, 117, 119, 121, 209, 212
- Todaro, S. A. et al. 76
- Trabasso, T. et al. 52, 74–6
- training: metacognition 213–14; metacognition for teachers 38–9; metacognitive knowledge monitoring 123; self-explanation in reading 66–70; self-regulated learning 310–12
- Traktext 164–5
- Transactional Strategies Instruction (TSI) 10, 34–5
- Triple-A-Challenge Gameshow 343, 344*f*
- Troia, G. A. 235
- Turkle, S. 356
- Turner, J. C. et al. 368–9, 434–5, 437
- tutoring: meta-knowledge 3–4, 361–76; computer tutors 367, 371–2, 376; discourse alignment 367*t*, 368; expert tutors 361, 370, 372, 373, 374, 375, 376; feedback 367, 367*t*, 373; flow 369, 375; grounding referents 366, 367*t*; hints 368; illusions 367*t*; knowledge transfer 367*t*, 374; meta-affect 368–9, 375–6; meta-knowledge of tutees 363–9; meta-knowledge of tutors 369–76; metacognition 365–6, 372–4; metacommunication 366–8, 374–5; non-expert tutors 361, 370–2, 373, 374; pedagogy 370–2; peer tutoring 361, 376; student mastery 367*t*, 373; tutoring frame 364; *see also* intelligent tutoring systems
- Valiante, G. 227
- Valli, L. 247
- Van Damme, J. 116, 123
- Van den Broek, P. 12
- Van der Meij, H. 48
- Van Merriënboer, J. J. G. 396
- VanLehn, K. 385–6, 397
- Vassileva, J. 393
- Vaughn, S. et al. 35
- Veenman, M. V. J. 211, 212
- verbal reports as data 7, 10
- Von Aster, M. 208–9
- Vygotsky, L. S. 7, 340
- Wallsten, T. S. 422, 426
- Walonoski, J. A. 397
- Weaver, C. A. et al. 86, 87, 93*t*, 94*t*, 101, 423, 424, 433–4, 435
- Web of Inquiry 189–94; Inquiry Cycle 179, 189, 190*f*; inquiry processes, goals, activities 189–90; managing inquiry 192; reflection 191, 191*f*; software advisors 190–1; evaluation 192–4, 193*t*
- Webb, T. L. 119
- Weerasinghe, A. 389
- Weinstein, C. E. et al. 99, 115
- Wharton-McDonald, R. et al. 245
- White, B. Y. 180, 187, 199
- Whitehouse, K. 283
- Wigfield, A. 301
- Wiley, J. et al. 87, 94*t*, 99
- Williams, J. P. et al. 36–7
- Wilson, R. D. 279
- Windschitl, M. 241, 247–8
- Winne, P. H. 261–2, 261*f*, 263, 264, 265, 268–9, 292, 322, 323, 330
- Winograd, P. 12
- Wise, A. 112
- Witte, S. 154, 156, 157, 159, 160
- Wittwer, J. 112
- Wong, B. Y. L. 210–11
- Wood, D. J. 393
- Wood, G. et al. 243
- Wood, H. A. 393
- Woolverton, M. 375
- working self 2
- Wright, G. N. et al. 233
- writing as applied metacognition 154–70; applied metacognition 160–1; definitions

- 154, 155, 160–1, 169–70; external symbolic representation 159–60; goal-directed metacognitive monitoring and control 157–9; investigative methodology 164–9, 166*t*, 167*f*; metacognitive model of writing 161–4, 162*f*; oneself or others 155–7; production of thought 155; revision 166, 169
- writing, children's 3, 131–50; areas of competence 131; cognitive processes 136, 137; Cognitive Strategy Instruction for Writing (CSIW) Program 142; knowledge telling 137, 138; knowledge transforming 138; and learning disabilities 138, 142; long-term memory 136, 137; metacognition: (knowledge about cognition 133–4; major elements in writing 134–6; regulation and control of cognitive activity 134); models of writing process 136–41: (Bereiter and Scardamalia 137–9; Hayes 137; Hayes and Flower 132, 133, 136–7, 140; Zimmerman and Reisemberg 139–40); novice and expert writers 137–8; primary students 31; process approach 132–3, 141; research 132; self-efficacy 135–6, 139; Self-Regulated Strategy Development (SRSD) 35, 142–50; self-regulation 135–6, 135*t*, 139, 140; socio-cultural theories 140; stages of writing 132; strategies instruction 141–2; task environment 136, 137; theoretical pragmatism 140–1; working memory 137
- Yanniv, I. et al. 226
- Yates, F. et al. 225*t*, 226, 231
- Yates, J. F. et al. 231
- Yuill, N. 50
- Zabrucky, K. M. 86
- Zelinsky, E. M. 92*t*
- Zimmerman, B. J. et al. 121, 135, 135*t*, 139–40, 185, 305, 306, 307, 308–10, 311