

school and for guidance in counseling students in their courses in graduate study (p. 28). There is considerable irony in the fact that most accredited graduate schools of psychology depend upon GRE test scores despite the fact that such scores have no acceptable proven or provable validity.

For a period of time, some test-maker bulletins omitted validity correlation statistics entirely. For reasons best known to the test-makers, validation information was for a time, not published in the test information sent to and read by the student. In order to obtain validation statistics, the bulletins instruct SAT student applicants, for example, to order a second manual called the ATP guide. The reference to this second guide is not prominent in the bulletin.

The 1987-88 ATP Guide admits that the SAT-verbal and mathematical predictive correlation is 27% for 10% of the colleges measured (.52x.52=27%), between 13% and 27% for 40 percent of the colleges (.36-.52), between 4% and 13% for 40 percent of the colleges (.21-.36), and below 4% for 10 percent of the colleges. ETS admits, "The validity of high school record is typically somewhat higher than the validity of the optimally weighted combination of SAT scores." ETS claims that the weighted combination of the high school record and SAT scores by a correlation addition of less than one half percent (9.07x.07). The ETS fails to state how the data should be weighted. There is no indication in the ATP Guide that any admissions director or admissions committee weights SAT scores or high school grades in the admissions process. (The College Board, 1987).

The 1984-85 Graduate Management Admission Test Bulletin of Information resolved validity disclosure problems by the simple expediency of not publishing validity information to test applicants. GMAT disclaimers are in comparison, much stronger than those provided with the GRE. ETS admits that the test, "cannot and does not measure all the qualities important for graduate study in management and other pursuits, whether in education, career, or other areas of experience; . . . (2) there are psychometric limitations to the test—for example, only score differences of certain magnitudes are reliable indicators of real differences in performance. Such limits should be taken into consideration as GMAT scores are used."

Employment test validity information provided by the ETS for tests such as the NTE teacher's test is also less than a resounding vote of confidence. The NTE teacher's test is sold to states and counties without validation. ETS simply tells prospective users to validate the test themselves. Incredibly, state after state has bought the test with that proviso.

The test-makers have not and cannot validate these tests with ethically applied, generally accepted statistical methods. A more serious question involves whether or not the test-makers use vague, ambiguous or highly technical disclosure information. The average applicant taking a predictive test is not skilled in statistics or psychometrics. Why then, do the test-makers persist in using statistical and psychometric language in

place of plain English? Why are correlation figures used in place of percentages? The answer may be that the plain English information does not look very good. The data provided by the test-makers constitutes prima facie proof that forcing students or job applicants to take predictive tests is an economic and human waste.

Why don't the test-makers and their affiliates publish percentage statistics? Would you publish percentage statistics if your correlations were this bad?

Practical Considerations

The actual field use of predictive tests is even more interesting than their statistical shortcomings. A large number of prospective law school applicants expressed concern when the 1982 LSAT test was announced, and they rushed to take the old test. The same thing happened in 1991 when the test was changed once again. As a result, applicants for the 1983-84 and 1992-93 school years are believed to be heavily represented by those who took the old test while applicants for the 1984-85 or 1993-94 school years are a mixed group. There is no ethical justification to support the use of two entirely different tests in selecting a particular law school class or any other class.

The Richardson School of Law at the University of Hawaii as but one example, admitted as much in a 1993 report to the Hawaii Legislature footnote (p. 12): "It is impossible to compare Law School Admission Test (LSAT) scores for all 20 years of the law school because both the test and the scoring system of the LSAT exam have changed during that period. The three different score ranges used since 1973 are not comparable. When the law school first opened in 1973, the range of scoring was 200-800; from the early 1980s until 1991, the test was scored on a 10-48 range. The latest scoring scheme—120-180—was first effective with the 1992 entering class." Here we have an accredited, ABA approved law school admitting LSAT scores over the years "are not comparable" and yet LSAT are still used to deny admission to applicants. In fact, either the new test score or the previous test scores were accepted for a time during an overlap period by educational institutions whenever new tests were introduced. That created a situation where a particular class would be entered using two different test score "schemes" despite the fact that they "are not comparable."

Another weakness with the practical use in the field of predictive test scores involves the limited psychometric background of those using the test scores. Most of those who make final selection decisions have no training whatsoever with regard to the limitations of predictive tests. Few decision makers understand the meaning of the psychometric cautions or the disclaimers found in testing literature. The average selection committee member may be reading far more into test scores than they should. To the extent that a situation has been created where users have too much confidence in the tests, the responsibility lies both with the test-maker and the institution requiring the tests. Additional responsibility lies with those in the academic community who know better and keep quiet.

At least one fully accredited, ABA approved law school, has a six person admissions committee two of whom are law students elected by the student body. Both are able to lobby and one has voting power. If any of the student admissions committee members have training in psychological testing, it would have to be by pure coincidence. When test makers send out test results, they routinely disclaim any responsibility with regard to the educational qualifications of those using their test results. The Standards for Educational and Psychological Tests and Ethical Principles of Psychologists of the American Psychological Association are simply ignored.

Admissions committee members may also be missing other important cautions found in standard psychometric texts such as Graham and Lilly's *Psychological Testing* (1984). Graham and Lilly caution (p. 42), "If not all people can be accepted by an institution, those admitted should be randomly selected in the absence of any validity information. Only if the test scores are not used in the selection process can an accurate determination of the predictive validity of a test be made." Once predictive test scores are used in the admissions process, any hope of determining validity based upon generally accepted statistical models is destroyed.

Graham and Lilly also note (p. 40), " * * * being able to predict who will be successful in a given job, whether as a police officer or airline pilot (or we might add, a physician, psychologist or an attorney), saves the person involved from an embarrassing failure and the institution from possible economic loss." The statement fails to deal with the embarrassing failure of not being admitted to graduate school. The statement also fails to deal with the potential economic loss to the applicant and the community despite the equal opportunity laws and constitutional protections of this country.

The uneasiness that continues to surface in the literature with regard to predictive tests (Fitzpatrick, 1983; Guion, 1978; Tenopir, 1977; Messick, 1980; Federal Trade Commission, 1978; Owen, 1985) comes from the knowledge that criterion information is far from perfect. It is well known that grades in graduate programs are a function of, and are influenced by, many factors other than academic aptitude. In the real world, criterion information represents a measure of convenience. There is no evidence that the criteria measured proves anything (Graham & Lilly, 1984).

The most important criterion from society's point of view is not grade point average, but the far more important criterion of excellence in one's chosen profession. The criterion actually used in this context is a compromise between one that is ideal and one that is readily available.

Substantial legal questions are involved whenever educational and employment tests are used in the admissions or employment process. Not only are careers being decided, the applicant is forced to pay for the privilege of taking a test that cannot be validated using either statistical or ethical principles. Those who make decisions utilizing predictive tests are vulnerable pursuant to federal and state privacy, due process, equal opportunity, and civil rights laws.