DOCUMENT RESUME

ED 171 800 TM 009 507

AUTHOR Ferguson, Richard L.; Brennan, Robert L.
TITLE Some Issues of Predictive Validity Related to College Admission Testing Programs.
PUB DATE Apr 79
NOTE 13p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Entrance Examinations; Grade Point Average; *Grade Prediction; Higher Education; *Predictive Ability (Testing); Predictor Variables; *Success Factors; *Testing Problems; Testing Programs; *Test Interpretation; Test Validity
IDENTIFIERS *American College Testing Program

ABSTRACT

Issues associated with the ACT (American College Testing) Assessment Program illustrate the need to assess college entrance examinations in terms of the variety of common predictive uses such as estimation of first year grade point average (GPA), or chances of obtaining a C average after certain semesters or in specific courses. Test developers should also be concerned with the diversity of groups and colleges to be tested. The problems professional schools face illustrate the need to examine the relationship between the predictor (test score) and the success factor (usually college GPA). Although tests are developed to spread students over a wide scale, professional school candidates score within a narrow range, thus affecting the predictive ability of the tests; the same narrow range problem applies to grades. Other prediction issues should be considered--hidden variables in college admissions, use of first-year GPA as the sole criterion, specification of the measure used to predict relationships; consequences of misclassification; test bias; and inability to measure nonacademic skills. In summary, a predictive decision is practical only when its various possible uses, misuses, and relationships to ultimate goals are considered. (CP)

SOME ISSUES OF PREDICTIVE VALIDITY RELATED TO COLLEGE ADMISSION TESTING PROGRAMS*

Richard L. Ferguson
Robert L. Brennan

The American College Testing Program

This paper is intended to address itself to a seemingly straightforward issue--namely, "Can the predictions of academic success currently achieved with national colleges admission testing programs be improved?" This question, however, is deceptively simple; and it begs for amplification and elaboration. Indeed, it seems to us that the question cannot be addressed meaningfully without a consideration of at least several issues pertinent to the validities of college admission testing programs, when data elements in those programs are used to predict academic attainment. In this presentation, therefore, we consider the prediction question in light of several general issues including: (1) objectives and purposes of prediction, (2) predictor and criteria considerations, (3) utility and prediction and (4) other considerations. Although these aspects of the prediction question are by no means pertinent only to testing related to college admissions concerns, this paper will concentrate solely on testing in that context.

Objective and Purposes of Predictors

The immediate challenge facing those who would seek to improve the criterion-related validities of college admissions tests, where

*This paper was presented at a symposium entitled "Can Predictions of Academic Success Be Improved?" at the Annual Meeting of the National Council on Measurement in Education, San Francisco, California, April 1979. The purpose of the paper was to stimulate discussion of several issues pertinent to the topic. Consequently, those issues and their consequences are not fully developed in the paper.

"academic success" is the criterion, is to further define the objective(s) of prediction. One might address the question "Can predictions of students' academic success in college be improved?" simply by investigating approaches to increasing the magnitude of a correlation coefficient relating test score performance to overall college grade-point average (a common approach in many prediction studies). Such an approach might be reasonable, but it is all too often rather simplistic, shortsighted, and misleading.

Although college admission tests are popularly viewed as having as their primary (if not single) purpose the prediction of student performance and the consequent selection of students, such views are not only greatly exaggerated, they are rather inaccurate. Even in those instances when predictive data based on tests are used in the selection process, test scores are seldom, if ever, used as the sole criterion for selection. Also, college admissions tests are used for a variety of purposes, many of which involve a prediction function and some of which do not. By way of illustration of the predictive uses, several examples of such uses are cited below following a general description of a college admissions program, the ACT Assessment Program.

The ACT Assessment Program is a multi-purpose assessment battery containing four ability tests, an interest inventory, and an approximately 200-item Student Profile Section aimed, in part, at describing prospective college students and their needs, independent of test scores. The four tests: English Usage, Math Usage, Social Studies Reading and Natural Sciences Reading, are each highly curriculum-related. The Program is supported with an array of technical and interpretive documents and

services including various research services used to assist colleges and students in predicting the academic success of students prior to enrollment in college.

One goal of college admissions testing programs is to provide colleges with information on the performances of present students which they can generalize to future students. Because colleges differ on many important characteristics, as do the groups of students who attend those colleges, the acquisition and use of such information is of necessity, an individualized matter for the college. Accordingly, ACT provides Research Services that enable individual colleges to consider the relationships between ACT test scores along with other pertinent data (e.g., high school grades and other test scores of the college's choice) and academic performance in college. Each year, ACT aids approximately 1,500 colleges in conducting regression studies aimed at validating each college's particular interpretations and uses of the ACT data. For example, in a typical prediction system the ACT tests are used jointly with high school grades (and in some instances with other data provided by colleges) to predict academic success in a number of contexts including estimating students':

(1) first-year (or first-term) college grade-point average

(2) chances of obtaining an overall grade-point average of C or higher at colleges specified by the students

(3) chances of obtaining an average first-year grade-point average of C or higher in particular schools or programs within a specific college or university

(4) chances of obtaining a grade of C or higher in specific courses offered by a college.

The above should serve to illustrate the need to view attempts to improve the prediction of academic success in terms of the variety of predictive uses commonly employed. Indeed, efforts to improve predictions of academic performance may meet with various degrees of success as a function of the particular predictive use being made of the test scores and other data in any given circumstance. Even if such consideration is given, it is perhaps pertinent to observe that college admissions testing programs are often used for purposes other than prediction, purposes which cannot be ignored as researchers attempt to identify procedures for improving predictions.

In addition to considering the particular uses of tests and other data for predicting students' academic performance, the test developer must also be concerned with the various types of colleges and groups of students to be tested and the validities of the tests and other data for predictions relevant to these colleges and groups. This is especially true of college admissions tests which are used with diverse types of colleges and equally diverse groups of students. It is possible, for example, that interventions aimed at improving the validity of the tests for predicting students' academic success could achieve the desired end for one group, but have a negative impact on another.

A particular use of the ACT tests illustrates this observation. Many of the colleges using the ACT tests to develop prediction equations find that the regression analyses conducted show that the Natural Science Reading Test usually does not add appreciably to the strength of the prediction of first-year college grade-point average. On the surface, it would seem that, for these colleges, the Natural Sciences Reading test might well be omitted from the ACT battery. To do so, however,

would be to reduce the effectiveness of the battery for the large number of nursing and engineering schools who find that this test is a major element in their prediction systems. Moreover, it might well be argued that, even if the ACT Natural Sciences Reading Tests failed to contribute in a significant way to predicting students' academic success at any college, the test should still be included if only because the battery was designed to assess knowledge and skills representative of the major secondary curriculum areas. To put it another way, predictive validity for a specific purpose is one, but only one, type of relationship in a network that addresses the construct validity of an instrument or program.

A corollary observation is apropos here. It must be noted that the objective of a person engaged in prediction is to investigate an interpretation of data arising from a specified measurement procedure. Moreover, as Cronbach has aptly stated, "Empirical validation is inevitably retrospective, yet the use of the test is always prospective."[1] For example, frequently a prediction system is developed using a group of examinees in a particular year, but decisions are ultimately based upon a different group of examinees tested in a subsequent year. Also, prediction of academic success using the data elements in college admissions testing programs, and efforts to sustain those improvements over time, are sometimes constrained by the fact that the predictor instruments/procedures are in a state of flux. Among other things, for example, test content in the ACT Assessment Program, is effected periodically by evolving curricula.

[1]Cronbach, L. J. Test Validation. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971, p. 485.

Moreover, the nature of the decisions made based on the available data (and the prediction systems that use those data) also change. Again, by way of example, ACT test data are clearly used much less frequently today in selection decisions than they were 15 or 20 years ago. Moreover, the data are used much more extensively for placement decisions than ever before, partly because of increased heterogeneity in the abilities of enrolled college students.

For the reasons cited above--multipurpose uses of test scores and other pertinent data involving prediction and the diversity of the groups which are the object of those predictions--the question "Can predictions of academic success in college be improved?" has no single simple answer that is entirely satisfactory or useful to those who routinely engage in such predictions.

Predictor and Criteria Considerations

In considering any single predictive validity related to a college admissions testing program, we must, of necessity, pay careful attention to the predictor(s), the criterion, and their relationship(s). A thorough treatment of these matters is beyond the scope of this presentation, but we will briefly consider some issues we judge important that are perhaps too frequently overlooked.

First, if we assume, as is usually the case, that the academic success criterion is college grade-point average, then we must also recognize that the nature of both the predictor variable(s) and the criterion variable(s) impact the relationship between the two. Moreover, test developers and researchers cannot easily impact some factors (e.g., practices related to grading) in the same manner that they can effect

improvements in test development procedures and techniques for analyzing the data resulting from the use of those procedures. The problems faced by many of the professions who use admissions tests in selecting first-year students for the professional schools are illustrative of the previous point.

The applicants to a typical professional school far exceed in number the available spaces. Moreover, these applicants are, as a group, usually extremely capable as demonstrated by excellent pre-professional school academic records, and equally high test scores on admissions examinations. The consequent restricted ranges of the predictor variable(s) alone impact the effectiveness of efforts to predict academic success in the professional schools. Although it is theoretically possible to extend the ranges of these variables, it is not always appropriate to do so. For example, although the developer of the admissions tests might construct those tests to spread students out across a test score scale more than the current tests do, it is likely that this could be accomplished only by drawing finer distinctions than are meaningful or appropriate given the goal of prediction--to select students who are capable of academic success in the professional school.

Also, it is frequently the case that the range of the success criterion is similarly constrained. Because the students admitted to a professional school typically have exceptional academic credentials, they usually perform extremely well in school. Indeed the attrition rate of many of the professional schools is very low, and seldom does attrition occur for solely academic reasons. A result is that the criterion for academic success is typically quite restricted in range. These facts, in part, no doubt explain the modest predictive relationships

so frequently obtained when undergraduate grade-point averages are combined with admissions test scores to predict professional school grade-point averages.

A second observation seems pertinent as one considers the objective of improving predictions of academic success in the context of admissions tests. Such tests are typically used on a continuing basis, year after year. In such cases, and when the tests (and other data) are used by schools that are highly selective of their students (an ever-diminishing number in these days of decline in enrollment), one might expect a decline in the magnitude of some types of predictive relationships of test scores (and other data) and academic success (college GPA), if the predictive system does indeed identify students who are likely to achieve academic success once admitted to college. It is perhaps interesting to note that data on a random sample of colleges who have been selective of students, and who use the ACT tests along with high school grades for predicting academic success, indicate no such discernible declines in multiple r's over the past several years. This may simply mean that the colleges are not relying solely on test scores (and other academic data related to prior performances of students) in selecting their students. To the extent that this is the case, and to the extent that other "hidden" variables are an integral part of the admissions decisions process, efforts to increase multiple r's (even for selective colleges) may have relatively little practical significance even if they are successful.

A consideration of the "success" criterion would not be complete if one failed to call attention to the shortcomings of using first-year college GPA as the sole criterion of prediction. Although it may be argued that academic success in the first year of college or a professional

school is highly relevant to subsequent academic success in college (since most attrition due to academic failures occurs in the first year of college or professional school), many also argue that grade-point-average, per se, is not a worthy success criterion. In that connection, many studies have been conducted to ascertain the relationship between achievement in college (and of test scores) and successful performance on the job. Most such studies have failed to reveal such relationships. One possible explanation for these disappointing results is that the predictor variables were too gross; that is, encompassed too many irrelevant and confounding factors.

Some recent attempts to relate new types of measures of academic performance in college to success in job performance have yielded some encouraging but very modest results. Notable among these is the College Outcome Measures Project (COMP) sponsored by ACT and funded, in large part, by the Fund for the Improvement of Postsecondary Education. This project, which focuses on the assessment of generic skills that represent, by the concensus of the colleges participating in the project, many of the intended outcomes of their general education or liberal arts programs, is in its third year of research and development. Preliminary studies suggest relationships between performance on various components of the COMP Measurement battery and successful performance in work directly corresponding to those components. Regardless of the ultimate success or failure of research like that underway in COMP, those engaged in the use of college admissions tests are well advised to continue to question the use of predicted college GPA as the sole criterion in selection decisions.

Finally, in the above discussion, the phrase, "predictive relationships," has frequently been used without specifying any measure of this relationship. All too often, such relationships are reported, interpreted, and/or discussed solely in terms of some type of sample correlation coefficient. To do so frequently obscures important facts, and possibly misleads decision-makers or others interested in interpreting a study of predictive validity. In this regard, at lease three somewhat inter-dependent points can be made. First, frequently the nature of the issues addressed by a predictive validity study suggest that a coefficient corrected for attenuation is more appropriate than a sample coefficient. Second, a correlation coefficient per se is seldom a sufficient basis for a practical decision--at a minimum, one should also consider score variances and measures of error, such as the standard error of estimate. Third, regression functions are almost always of more fundamental importance than the magnitudes of sample correlation coefficients. These points have been made by many writers, but they are sufficiently important to justify repetition.

Utility and Prediction

Predictions are properly made only in conjunction with the various potential consequences of each predictive option. The concern of the test developer and the test user should not be focused solely on minimizing the number of classification errors, per se, but also on the consequences of misclassifications. To the point, it is desirable to link questions of prediction with concern for utility or loss functions.

Given tests like those contained in the ACT and other admissions testing programs (SAT, LSAT, MCAT, etc.) and the diverse types of decisions

made using data from these tests, (often in conjunction with other data), the particular loss functions of interest vary widely. In the case of the ACT tests, for example, the consequences of classification errors could be markedly different when the tests are used in admission decisions by highly selective postsecondary institutions as opposed to open admissions institutions, the latter being far more frequent than the former at this time. In the latter case, the uses of the ACT tests with students typically include academic advising and placement. In such uses, students are frequently able to accept or ignore the counsel that the test data provide.

Given the merits of the above observations, a reasoned decision of a predictive nature requires, in principal, a joint consideration of both utilities (or values) and statistical relationships of scores. It is necessary to acknowledge, however, that practical problems remain even if the user of test data (and/or other data frequently used in predicting academic success) combines a concern for the expected consequences of decisions based on predictions with a parallel concern for the adequacy of the statistical procedures used to set the stage for those decisions. For example, college admissions testing programs are used in a wide variety of contexts for various different purposes; and utility or loss functions are specific to both context and purpose, at least. However, it is generally not a simple matter for a particular decision-maker in a particular college to specify and use losses or utilities for a particular decision. It is to be hoped that the decision-maker's task will be facilitated by mechanisms such as the Computer-Assisted Data Analysis System developed by Novick and his colleagues.

Other Considerations

In those instances where test and other data are used principally for selection decisions (perhaps most frequently in the case of admissions testing for professional schools), there are at least two very important additional considerations. First, selection procedures based on grade predictions for all students may be unfair in some ways to certain groups (e.g., those of minority racial-ethnic background or of atypical age). When an institution is concerned only with maximizing the chances of success of selected students, the use of a single predictive systems may be appropriate. However, when the institution's goals include concern for a kind of "fair" opportunity for admission for all subgroups within the applicant population, then other procedures may be required.

A second consideration about the sole use of predicted GPAs for selective admissions is the effects of that practice in limiting the selected group to primarily academic achievers to the exclusion of other types of achievers. Even selective institutions may wish to provide opportunities for the slightly less academically talented student with special talents in leadership, music, art, writing or other such areas.

Summary

In conclusion, the general goal of improving predictions of academic success in college is inherently worthwhile, as one aspect of a continuing effort to enhance a particular type of validity generally appropriate for a college admissions testing program. That goal takes on practical significance, however, only when it is considered in the context of the various possible and likely uses of the predictions, the utilities associated with those uses, and the relationships of those uses to the ultimate goals of the educators who rely on them.