Nader's Raid on the Testing Industry

Is It in the Best Interest of the Consumer?

Robert M. Kaplan
Nader’s Raid on the Testing Industry

Is It in the Best Interest of the Consumer?

ROBERT M. KAPLAN   San Diego State University

ABSTRACT: Recent attacks on the use of the Scholastic Aptitude Test (SAT) are considered in light of their expected impacts on consumers. The widely publicized argument that the SAT is no better than chance in predicting college performance is based on a misunderstanding of basic statistics. Extensive evidence suggests that the SAT, in combination with high school grades, is a very good predictor of college success for students from different ethnic groups and income levels. The suggestion that the use of SAT scores is a ploy by the rich to deny opportunity to low-income students is not well supported by a variety of validity studies. Although legislation stemming from the Nader investigation of the Educational Testing Service purports to protect consumers, it may actually work against them. As a result of the legislation, consumers can expect to pay a higher price to be evaluated by a less valid procedure.

Public controversy is not new to the field of psychological testing. As Cronbach (1975) noted in a much-acclaimed paper, mental testing has been scrutinized by the American public since its large-scale introduction more than six decades ago. Throughout the years, public concern about tests has forced psychologists to continually reevaluate their methods, and this process has stimulated the continued development of more scientifically justifiable tests. Despite these improvements in the science of testing, psychological tests have recently come under new attacks. Among the most harsh but least justified attacks has been the review of the Educational Testing Service (ETS) by Allan Nairn and associates (1980a, 1980b) under the auspices of Ralph Nader’s consumer advocacy group. Although the report was authored by Nairn, it has come to be known as the Nader or the Nairn/Nader report.

The Nader Raid

Ralph Nader and his associates have won considerable public acclaim for their attacks on large corporations and their advocacy of the consumer position. Recently the group became interested in the nonprofit Educational Testing Service (ETS). ETS was created by the College Entrance Examination Board (CEEB), the American Council on Education, and the Carnegie Foundation in 1948. Their original and best-known mission was to create and administer aptitude tests such as the Scholastic Aptitude Test (SAT). By 1979, ETS was responsible for over 300 testing programs, including the Graduate Management Admission Test (GMAT), the Graduate Record Exam (GRE), the Multi-State Bar Exam, and the Law School Admissions Test (LSAT).

In a much-publicized series of attacks by Nairn and other consumer advocates, ETS has been characterized as an evil bureaucracy operating under a guise of secrecy to maintain the status quo. The critics suggest that the ETS testing programs, in particular the admissions testing program of the College Board, perpetuate the use of invalid and biased tests to assure that only those from wealthy families are admitted to the most prestigious colleges and universities. Throughout the report, Nairn et al. (1980b) refer to well-known data or comments that support the ETS position as “commissioned by ETS” or “prepared by ETS staff.” This tactic attempts to discredit work done by well-respected ETS scientists as biased and not to be believed. Before continuing, it is worth noting that I have no connection to ETS nor do I have any reason to be under their influence. Some of the data presented in this paper are taken from ETS
documents, yet these are the same data used by ETS foes.

Two arguments underscore the Nader group’s attack on ETS. The first is that the Scholastic Aptitude Test (SAT), which is the major instrument of the College Board’s admissions testing program, is not a valid predictor of college success. The second argument is that the SAT scores reflect family income more than they do scholastic potential. I will examine each of these criticisms briefly.

Reliability and Validity of the SAT

In comparison with most psychological tests, the psychometric properties of the SAT are quite impressive. Each of the 85 verbal (SAT–V) and 60 mathematical (SAT–M) items is a five-option, multiple-choice task scored by a formula that is intended to offset any gain in score that might be expected from blind guessing. Test–retest, internal consistency, and alternative form reliability coefficients consistently range from the high .80s to the low .90s. Recent documents suggest that the KR–20 reliability of the SAT is usually found to be around .91 for the verbal section and about .92 for the mathematical section (College Entrance Examination Board, 1979). These high reliability coefficients suggest that SAT scores are internally consistent and reproducible using different forms of the test.

In addition to providing a good record of reliability, much work has gone into validity documentation. Validity defines the appropriateness of inferences based on a score or measure (Kaplan & Saccuzzo, 1982). Most inferences based on SAT are predictive. Predictive validities are the correlations between the test and performance on some well-defined future criterion, usually first-year college grade point averages. Since college grades cannot be known when students are still in high school, an SAT score and knowledge of the test’s predictive validity can help college admissions officers to forecast success. A variety of studies summarized by ETS shows that predictive validity coefficients for freshman grades vary widely depending on major, with the lowest coefficients being in the teens and the highest coefficients being in the low to mid .60s. Studies using students in liberal arts and general programs showed that median validity coefficients for predicting first-year college grade point average were relatively good; the coefficients for each portion of the test were in the mid .40s for women and in the mid .30s for men. The coefficients were considerably higher for students scoring in the top decile of the SAT. Although high school grade point average (GPA) was a better single predictor of college success ($r = .48$ for women and $.45$ for men), college performance was best explained by the linear combination of SAT–V, SAT–M, and high school GPA. Using data from the College Entrance Examination Board (1979), the linear combination explained relatively well college performance ($R = .54$ for men and .61 for women). In general, the psychometric soundness of the SAT is based on years of experience, statistical analysis of results from tens of thousands of administrations, and careful item selection. What then is the concern?

In their attack on the use of the SAT, the Nader group held ETS accountable for “perfect” prediction. They asserted that “ETS statements to students about the predictive accuracy of its tests are full of the word ‘perfect’ . . .” (Nairn et al., 1980b, p. 59). A review of the current SAT statement to students revealed this statement to be untrue. Instead, the ETS statement appropriately acknowledges that a perfect tapping of true abilities cannot be accomplished with the SAT. It then states, “The precision of any test is limited because it represents only a sample of all the possible questions that could be asked, and because people perform at different levels at different times for reasons unrelated to the characteristics of the test itself” (CEEB, 1979, p. 9).

The Nader report continues by reporting data analyses that demonstrate only marginal understanding of statistical procedures and psychometrics. For example, the report seems to confuse the concepts of reliability and validity. First it comments on the ETS discussion of the standard error of measurement, which correctly informs students that on the basis of reliability studies, the chances are about two in three that their observed score falls within one standard error of measurement around the true score.

In the next paragraph, the report switches the focus from reliability to validity without informing the reader of the distinction between the two concepts. It suggests that the SAT correctly predicts college performance in only about 11.9% of the cases and that this percentage is just slightly better than chance. In a footnote separated from these passages by nearly 400 pages, the report acknowledges that recent ETS documents emphasize that the two-thirds figure refers to the test’s reliability and not its validity. Yet in the text of the report, this point remains obscure.
SAT VERSUS CHANCE

It is instructive to trace the steps which led Nairn to conclude that SAT scores are no better than chance for predicting first-year college performance. First the report seems to confuse percentage of variance accounted for with percentage of accurate predictions. For example, some coefficients of determination (squared correlation between SAT and college performance) for SAT validity suggest that the test accounts for about 12% of the variance of first-year college performance (CEEB, 1979). Nairn et al. would interpret this to mean that the test only made accurate predictions in about 12% of the cases. The two concepts are simply not comparable.

The 11.9% accurate prediction figure also seems suspect since most validity studies report approximately a .40 correlation between SAT scores and first-year college grades. Thus, squaring the validity gives a coefficient of determination of .16, and multiplying by 100 (as Nairn does), 16%. The discrepancy between the 16% and 11.9% figures is rooted in another misunderstanding of elementary psychometrics.

Nairn arrived at the 11.9 figure through the following maneuvers. First he obtained data from an ETS report by Ford and Campos (1977) that listed combined-sex SAT median validity coefficients (with college success as the criterion) for each year between 1964 and 1974. Ford and Campos had obtained separate estimates of SAT-V and SAT-M validities for each year by calculating medians over many studies. The results of more than 2,000 validity studies were included in the Ford–Campos tables.

Nairn averaged the median coefficients across the 11 years (1964 to 1974) separately for the two components of the test to obtain a mean validity of .37 for SAT-V and a mean coefficient of .32 for SAT-M. He then averaged these two means to obtain .345. Next, he calculated the coefficient of determination by squaring .345 in order to obtain .119. Finally, he multiplied by 100 to obtain 11.9%.

At each step in this process, Nairn failed to apply widely accepted statistical methods appropriately. Correlations should not be averaged because they are not uniformly distributed across the -1.0 to +1.0 interval. Most textbooks suggest that they be converted to another metric (such as Z'), averaged, and then transformed back to correlations (Hays, 1968). A more serious problem is the averaging of the SAT-V and the SAT-M scores to obtain an overall validity. Since each portion of the test makes some independent contribution to prediction, the validity of the whole test is greater than that of either part. There would be no purpose in using the whole SAT if superior predictions could be made from one half of the test. The same report from which Nairn obtained his data shows the combined validity to be .41.

PERFECT PREDICTION?
The most serious mistake is equating the squared validity coefficient with the percentage of cases in which a test provides a "perfect prediction." This confusion is clear in the Nairn/Nader (1980b) report, which states that the SAT "... delivers an average accuracy—what statisticians call 'percentage of variance accounted for' or 'percentage of perfect prediction'—of 11.9 percent" (p. 60, emphasis added).

This error is then compounded by another misunderstanding of basic statistics. Nairn et al. argue that since only 11.9% of the cases can be predicted from the test, college performance might be predicted as well by chance. They argue that "to see how often chance would predict grade ranking within a group as well as the ETS aptitude test one subtracts the test's 'percentage of perfect prediction' from 100 percent" (p. 64). Using this logic, Nairn then suggests that the percentage of cases in which a random process would have predicted performance as well as the SAT is 100 - 12 = 88%. The error in this reasoning is that the percentage of variance not explained is not the same as the percentage of cases in which random predictions will be as accurate as those based on test results.

The conclusion that the SAT does only slightly better than chance is simply in error. Even the lowest reported validity coefficients are statistically significant at the .01 level. This suggests that the probability of obtaining the reported relationships by chance would be less than 1 in 100. Achieving this level of prediction by chance alone would be quite improbable. According to a recent ETS paper, the percentage of cases in which the SAT would be expected to make more accurate predictions than a random process should be 100% (Educational Testing Service, 1980b). The ETS claim can be defended using statistical theory, while there appears to be no theoretical or empirical justification for the Nairn/Nader position.

Tables taken from Minium's (1970) Statistical Reasoning textbook help to illustrate this point. Table 1 shows improvements over chance prediction that would be expected for tests having valid-

---

3
ity coefficients similar to those reported for the SAT. The first two columns list validity (test-criterion correlation) coefficients and coefficients of determination. Assume that the goal is to predict which students will be in the top half and which students will be in the bottom half of their college class. By coin toss, we would expect to be correct in 50% of the cases. The third column in Table 1 gives the proportion of cases in which the test would yield an accurate prediction in excess of the 50% expected by chance. The final column lists the proportion of improvement in prediction relative to chance prediction. It is obtained by dividing the value in the third column by the probability of obtaining a correct classification by chance (.50 in this example). For estimating which students will be above the median in first-year college performance, a test with validity of about .40 (such as the SAT) should increase accurate predictions by 13% beyond the chance rate of 50%. Further, it should provide a 26% proportional improvement over chance. A random selection procedure has a validity of .00 and would provide no improvement over chance. Thus, the Nairn statement that the SAT does no better than chance is without merit.

Another issue concerning the validity studies is the use of college grades as the criterion against which the SAT is validated. Although it is frequently argued that college grades are the ultimate criterion against which to evaluate any predictor, it is rarely acknowledged that college grades are themselves a very fallible criterion. Goldman and his colleagues have very convincingly demonstrated that grading standards vary greatly across different college majors. Thus, correlations with GPA are attenuated because the criterion is a composite of many nonequivalent components (Goldman & Hewitt, 1976; Goldman, Schmidt, Hewitt, & Fisher, 1974). One study conducted within the University of California system demonstrated that the validity of the SAT increases when the test is evaluated against the grading standards of individual professors (Goldman & Slaughter, 1976). These findings suggest that the SAT may be a better predictor of college performance when college performance is measured within homogeneous categories of academic work than it is when reported by ETS summary statistics.

The Nader group’s conclusions have now been widely disseminated through popular media (Nairn & Associates, 1980b). These incorrect conclusions could have been avoided by any student who had successfully completed a single course in testing or correlation.

The SAT–Income Connection

Another major argument in the Nader group’s attack on ETS was that family income is correlated with SAT scores. The rationale for current college admissions programs is that students are evaluated on the basis of merit rather than on the basis of economic background. This contrasts with earlier times when only the wealthy and the privileged could gain entrance to major universities. The College Board’s admissions testing program contends that SAT scores represent merit rather than family background. In contrast, the Nader group maintains that SAT scores primarily reflect social class and that the use of the scores by college admissions committees allows them to perpetuate an aristocracy by selecting high-income students under the guise of merit.

Although the argument that SAT scores reflect social class more than they do merit is interesting and politically appealing, the evidence for this position is not as strong as the Nader report suggests. Their case rests on a College Entrance Examination Board’s (1973) report on college bound seniors that shows mean family income broken down by SAT performance in 50-point intervals. The relationship is indeed linear and impressive. However, by presenting means, all of the individual variability is stripped away. The correlation between mean income and level of SAT performance is .965. However, the appropriate index is the correlation at the individual level. For the data used by Nairn, which include observations for 646,956 students, this value is .23 (CEEB, 1973). A more recent report based on a more refined measure of family income reports the correlation as .29 (Educational Testing Service, 1980a).

---

**TABLE 1**

*Value of Tests With Differing Validities for Improving Predictions Beyond .50 Expected Chance Level*

<table>
<thead>
<tr>
<th>Validity coefficient $r_v$</th>
<th>Coefficient of determination $r_2$</th>
<th>Correct predictions beyond $.5$</th>
<th>Improvement relative to chance</th>
</tr>
</thead>
<tbody>
<tr>
<td>.50</td>
<td>.25</td>
<td>.17</td>
<td>.33</td>
</tr>
<tr>
<td>.40</td>
<td>.16</td>
<td>.13</td>
<td>.26</td>
</tr>
<tr>
<td>.30</td>
<td>.09</td>
<td>.10</td>
<td>.20</td>
</tr>
<tr>
<td>.20</td>
<td>.04</td>
<td>.06</td>
<td>.12</td>
</tr>
<tr>
<td>.10</td>
<td>.01</td>
<td>.03</td>
<td>.06</td>
</tr>
<tr>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Note.* Adapted from Minium (1970, p. 199).
The use of grouped data in the Nader report also masked the fact that there is wide variation in income at each SAT performance level. Students at each income level obtain the full range of SAT scores. Nearly a third of the students from families with yearly incomes less than $6,000 (1974 dollars) obtained SAT scores that were above the median. Historical evidence indicates, contrary to Nairn's conclusions, that the use of standardized tests has resulted in increased acceptance of low-income students by selective universities. Low-income students with high SAT scores have been able to demonstrate their academic potential despite their attendance at less prestigious prep schools (ETS, 1980a).

There is no evidence that the SAT-GPA correlation is significantly lower than the SAT-income relationship. If the Nader group had used the same method of displaying the SAT-GPA relationship, they would have found an equally impressive linear relationship between average SAT scores (in 50-point intervals) and mean GPA. I calculated this correlation based upon 156 validity studies in 1974 (see ETS, 1980b, p. 18) and found the coefficient to be .999. Instead of reporting the GPA-SAT relationship the same way they had reported the income-SAT relationship, the Nader report adopted a misleading tactic. Using individual data for the GPA-SAT correlation, they included the error variance for the relationship they wanted to show as weak. However, by grouping the data for the income-SAT relationship, they excluded the error variance for the relationship they wanted to show as strong. When either individual or grouped (mean) correlations are used consistently, the SAT-GPA relationship emerges as the stronger correlation.

It is not surprising that SAT and family income are correlated. In fact, this appears to be one of the few points on which nativists and empiricists agree. Yet, the Nader report attempts to bolster the argument with data suggesting that first-year college grades do not correlate with income (p. 206). Actually this observation weakens rather than strengthens its argument. Figure 1 uses Venn diagrams to illustrate the set of relationships found to be disturbing to the Nader group. The figure shows the association between SAT and income as well as the relationship between SAT and GPA. In addition, it shows the independence of GPA and income. Inspection of the figure suggests that the SAT measures some aspect of GPA that is independent of the relationship between SAT and income. If this figure represented the true situation, it could be argued that the SAT is measuring both merit and social class. Yet the SAT-income relation does not enter the prediction of college success. In other words, SAT may measure the part of potential that is independent of family background. This position is further supported by the multitude of studies showing that the SAT has approximately equal predictive validity for differing income and ethnic groups (Cleary, 1968; Kallingal, 1971; Pheifer & Sedlacek, 1971; Sattler, 1982; Temp, 1971). Although the Nairn/Nader (1980b) report reviews studies on test bias in some detail, it apparently finds studies showing equal validity unconvincing. The authors' reasons are reflected in their beliefs about the validity of the SAT and the family income-SAT connection. With regard to the data on equal validity for different groups, they state, "... when those scores bear an essentially random relationship to ability to succeed in one's chosen career, members of low-scoring groups are excluded for a reason unrelated to their actual potential for accomplishment" (p. 10). As I noted earlier, the conclusion that scores are related to future accomplishment in a random manner demonstrates a misunderstanding of basic statistics.

The connection between bias and family income is reflected in the following statement, "The lower average scores of minority applicants are primarily a reflection of what is perhaps the single most important characteristic of the test, its tendency to

![Figure 1. Hypothetical relationship between grade point average, SAT score, and income as suggested by the Nairn/Nader report.](image)

1 Although Nairn et al. (1980b) report that there is no correlation between family income and college grades (p. 206), White (1976) averaged 41 correlations from separate studies to obtain a correlation of .24 between school grades and socioeconomic status. Taking this into consideration, the family income and GPA circles in Figure 1 might overlap. Figure 1 represents the situation as suggested by Nairn. The argument begins with Nairn's interpretation.
rank people by income” (Nairn et al., 1980b, p. 113). As we have seen, there is a weak association between income and SAT performance and this relationship is confounded with race. However, the conclusion that a test is biased because members of one group earn lower average scores endorses only one among several alternative definitions of test bias (Flaugher, 1978). Psychometricians have devoted considerable thought and attention to the definition of test bias (Cleary, Humphreys, Kendrick, & Wesman, 1975; Dunnette & Borman, 1979; Flaugher, 1978; Jensen, 1980), and it has been convincingly argued that different definitions of test bias reflect different philosophical positions with regard to fairness. The Nairn/Nader report takes the position that the use of any test which results in disproportionate selection of non-minority students is biased. This endorses one of the philosophical positions discussed in the literature they reviewed (Hunter & Schmidt, 1976) without acknowledging the rationale for and consequences associated with the other positions. The position Nairn advocates may lead to greater numbers of minority students being accepted to college, yet it may also be associated with lower average performance among college students and a higher rate of failure among minority students (Dunnette & Borman, 1979).

It is simply unfair to characterize the field of psychology as unconcerned about test bias. In fact, considerable effort and thought have been devoted to the development of fair assessment procedures. The Nairn/Nader report presents one perspective, but the reader is advised to review scholarly publications that consider this problem from all perspectives (see Clarizio, 1979; Flaugher, 1978; Guthrie, 1976; Hunter & Schmidt, 1976; Jensen, 1980; Mercer, 1979; Sattler, 1982).

How About Unfair Selection?

It is possible that colleges and universities have used SAT results to make unfair and unwise decisions. The Nairn/Nader report correctly condemns college selection officers who rely exclusively on the SAT to select students. If the SAT accounts for only 12%-20% of the variance in college performance, it is clear that other variables must be considered. Some college admissions officers may incorrectly conclude that the SAT is a more powerful predictor than is actually the case. As unfortunate as it is that some test consumers are ignorant with regard to psychometrics, the blame does not belong exclusively to ETS as is implied by the Nairn/Nader report.

The SAT manual gives very clear advice on the interpretation of SAT scores (CEEB, 1979). The manual appears to meet and even exceed the criteria for educational and psychological tests set forth by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education (1974). Contrary to the implications of the Nairn/Nader report, the College Board specifically advises universities and colleges to consider SAT scores as one among many predictors of academic success. School administrators are specifically directed to consider high school grade point average, and the SAT form now includes a section on activities, interests, and awards in order to provide the schools with more information about the applicant.

The Nader group's assertion that college administrators should consider information in addition to SAT scores is almost gratuitous. I wholeheartedly support this and so does ETS. Surely we should not blame ETS for the abuses of those administrators who fail to read or comprehend the ETS manual.

If one purpose of the Nader group's report was to discourage universities from using only SAT scores for admissions decisions, they might find that some universities need not be persuaded. Times have changed and universities are making extensive efforts to recruit students. Nearly all high school seniors are admitted to at least one college if they apply, and 95% are admitted to either their first or second choice (Astin, King, & Richardson, 1978). Schools seem to differ greatly in what they look for, but few admit to using SAT scores exclusively. This is revealed in a recent series of interviews by reporters for Newsweek magazine. The magazine reports, “At mathminded Caltech, College Board scores are given the most weight because they seem so precise, other schools such as Atlanta's Emory rely more on high school grades. Wellesley seeks to create a 'microcosm of the world, with as much diversity as possible' within the strict limits of accepting only women” (Adler, Huck, Lee, & Abramson, 1980, p. 112). A recent survey of college registrars and admissions officers found that the SAT was cited as the "most important" factor for admitting students by less than 2% of selective universities (VanDusen, Nelson, Jacobsen, & Ivens, 1979). By these accounts the
suggestion that universities should use information other than the SAT would result in no change in current practice.

Some Consequences of Following Nader's Advice

Among the recommendations of the Nader group's report is that high school grade point average be used instead of SAT in the selection of college students. The rationale for this is that the zero-order correlation of high school GPA with first-year-college GPA is higher \( r = .49 \) than the corresponding correlation of either the SAT-V \( r = .38 \) or the SAT-M \( r = .37 \). Nevertheless, the exclusive use of high school GPA might not be advisable because studies have shown that the SAT provides an increment in the ability to predict college success beyond the information given by high school GPA. The multiple correlation using both high school GPA and SAT jumps to .56 (data from CEEB, 1979). Thus, it appears that the SAT has considerable validity beyond the information given by high school GPA.

In order to evaluate how much information is provided by the test, a simple decision model may be used. This model is based on a set of tables developed by Taylor and Russell (1939). The tables can be used to evaluate the validity of a test in relation to the amount of information it contributes beyond what would be expected without the use of a test. The percentage of persons who would succeed if they were not selected on the basis of a test is called the base rate. Table 2 is based on information in a Taylor-Russell table for a base rate of 60%. For this hypothetical example, success is defined as obtaining a first-year-college grade point average of 2.3 or higher. In a group of unselected students, 6 in each 10 would be expected to meet this criterion of success. Table 2 shows how we can improve on this ratio by selecting students on the basis of different criteria. Selective institutions that admit only about 20% of the applicants and which use only the SAT-M to select students would find that 79% of those selected would succeed. Among students selected by this institution using the full SAT plus grade point average, 88% would be expected to succeed. Institutions that are less selective obtain less benefit from the use of tests because they come closer to admitting all applicants. It is interesting that the Nairn/Nader (1980b) report acknowledges the value of Taylor-Russell tables (pp. 419–420) but does not connect this approach to their interpretation of validity coefficients.

Validity studies tend to show that the SAT provides some useful information beyond what is obtained from high school grades alone. Psychometric theory would suggest this because the tests in addition to grades and other information provide a better sampling of the skills necessary to succeed in college. The exact amount of additional information the SAT provides is difficult to estimate and varies with situation and method of analysis. The Nairn/Nader (1980b) report estimates the unique contribution of the SAT to be 5% of “perfect prediction,” and their regard for this increment is indicated in the title to their Chapter 5, which they call, “Five Percent of Nothing: Aptitude Testing, the Respectable Fraud.” In one of their tables, “predictive efficiency gained” is displayed for 12 separate years, yet the report fails to mention how these values were calculated. An ETS document concluded that the data were calculated using an index of forecasting efficiency \( (1 - \sqrt{1 - r^2}) \) which, for 1974 data, would give an index of .134 for high school grades and .185 for grades plus SAT scores. If these figures are converted into percentages by multiplying by 100, the difference between them comes to about 5%. However, to calculate the index, it is necessary to divide by the efficiency of grades alone (13.4% in this example). Unfortunately, Nairn forgot to do this. If he had, he would have needed to rename his chapter, “Thirty-Eight Percent of Nothing . . .” (ETS, 1980b).

There is little doubt that some students who

<table>
<thead>
<tr>
<th>Selection procedure</th>
<th>Approximate validity*</th>
<th>20%</th>
<th>50%</th>
<th>80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>.00</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>SAT-M</td>
<td>.37</td>
<td>79</td>
<td>72</td>
<td>65</td>
</tr>
<tr>
<td>SAT-V</td>
<td>.38</td>
<td>80</td>
<td>72</td>
<td>65</td>
</tr>
<tr>
<td>SAT-V and SAT-M</td>
<td>.41</td>
<td>81</td>
<td>73</td>
<td>66</td>
</tr>
<tr>
<td>High school GPA</td>
<td>.49</td>
<td>86</td>
<td>76</td>
<td>67</td>
</tr>
<tr>
<td>SAT and GPA</td>
<td>.56</td>
<td>88</td>
<td>78</td>
<td>68</td>
</tr>
</tbody>
</table>

Note. SAT-V is Scholastic Aptitude Verbal; SAT-M is Scholastic Aptitude Math; GPA is Grade Point Average. Criterion for success is GPA ≥ 2.3 out of 4.0.

* Based on a Taylor–Russell (1939) table with a base rate of .60. The validities are based on median values from a summary of 158 studies (CEEB, 1979).
TABLE 3

Advantages of a Selection Procedure

<table>
<thead>
<tr>
<th>Performance</th>
<th>Accept</th>
<th>Reject</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>18</td>
<td>42</td>
<td>60</td>
</tr>
<tr>
<td>Failure</td>
<td>2</td>
<td>38</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>80</td>
<td>100</td>
</tr>
</tbody>
</table>

Note. Rounded to nearest whole applicant. Based on a Taylor-Russell (1969) table; base rate = .60, selection ratio = .20, validity = .56. Probability of success given acceptance is 18/20 = .90; probability of success given rejection is 42/80 = .53.

would have succeeded in exclusive colleges are rejected on the basis of SAT scores. However, the use of valid aptitude tests in combination with other information attempts to minimize this possibility. Table 3 displays hypothetical data on the expected success of a cohort of 100 students who were selected by the SAT for a college that admits 20% of its applicants. Success, again, is defined as achieving a GPA of 2.3 or higher, and we would expect 60% of the applicants to achieve this criterion without a screening method. The analysis is based on a Taylor–Russell table for a base rate of .60, a selection ratio of .20, and the reported validity of the selection strategy as .56. This validity was obtained from the SAT manual as the median multiple correlation between the combination of SAT and GPA with first-year college success (CEEB, 1979).

Table 3 shows that the probability of success given use of the screening strategy rounds to .90, whereas the probability of success given rejection is only .53. Clearly, use of test scores and grades to select students excludes some talented people and allows some less capable people to be admitted. However, as the analysis shows, this procedure does maximize the likelihood of success. If we abandon this approach, as the Nader report implies we should, we can expect a greater rate of failure or a lower rate of performance in prestigious American universities.

Conclusion

The Nader group's report on testing grew out of legitimate concern for just treatment for American citizens. Clearly tests have been misused and the results of aptitude tests have caused serious discouragement and self-questioning for many individuals. Psychometricians and test constructors have an ethical responsibility to develop fair selection devices, and with the help of these devices, college administrators have the responsibility to assure that the available spaces are filled in a fair and just manner.

Although the Nader group appears to share this concern, it is likely that their actions will make the problem worse rather than better. Abandoning aptitude testing programs may provide access to selective colleges for some students who would have previously been rejected. Yet it will also mean that some students who would now be admitted will be rejected. Since the present system selects students with the greatest likelihood of success, abandonment of this approach seems ill advised.

In the absence of this system, students who have a high probability of success will be rejected in favor of accepting students with a lower probability of success. Nader's evidence that the SAT is not a valid predictor of college success is based on a misunderstanding of basic statistics. Nevertheless, it has been widely disseminated by the popular press, and it has had more impact than could have been expected if it had entered through a carefully refereed academic journal.

The truth-in-testing legislation that is advocated by the Nader group would force testing companies to make their items public. This may require continual redevelopment of test items without sufficient opportunity for validation. The impact of the Nader proposal could only serve to decrease the validity of the tests. In other words, the quality of the product may suffer. At the same time as quality would decrease, costs would be expected to increase due to the expense of continually creating and testing new items. Students might expect to pay a higher fee to take a less valid test. And this is being accomplished in the name of advocacy for the consumer!

REFERENCES


Cleary, T. A., Humphreys, L. G., Kendrick, S. A., & Wesman,


Goldman, R. D., & Slaughter, R. E. Why college grade point average is difficult to predict. Journal of Educational Psychology, 1976, 68, 9-14.

Guthrie, R. V. Even the rat was white: A historical view of psychology. New York: Harper & Row, 1976.


Nairn, A., & Associates. The reign of ETS. Today’s Education, 1980, 69, 586-646. (a)


