

Validity versus Utility of Mental Tests: Example of the SAT

LINDA S. GOTTFREDSON

Counseling and Personnel Services, University of Maryland, College Park

AND

JAMES CROUSE

Department of Educational Studies, University of Delaware

Scientific debate over mental tests has focused in the past on their validity, but recent debate has shifted toward questions of their practical value. Research by James Crouse is an example of this trend. Crouse has provided persuasive evidence that aptitude and achievement tests are valid for predicting later educational and economic success. However, he has also provided evidence challenging the Educational Testing Service's claim that college admissions decisions are improved by using the Scholastic Aptitude Test (SAT) to supplement the high school record. Major issues in the debate over Crouse's research and recommendations concern (a) the SAT's incremental value in admissions, (b) the nature of the admissions process, (c) other benefits to colleges of the SAT, (d) its possible benefits for college applicants, and (e) the use of college admissions tests as interventions for improving secondary education. Suggestions are provided for encouraging more analytical assessments of test utility. © 1986 Academic Press, Inc.

In his presentation to the Personnel Testing Conference of Southern California, James Crouse summarized his recent work concerning the costs and benefits of using aptitude and achievement tests in college

Because of prior commitments, James Crouse was unable to prepare an independent contribution to this special issue, "The *g* Factor in Employment." However, he kindly assisted in preparing this paper based on his talk, "Does the SAT Predict Academic Success?" that he presented at the 1985 fall conference of the Southern California Personnel Testing Conference. The issues discussed here receive much more extensive consideration and documentation in his forthcoming book with Dale Trusheim (Crouse & Trusheim, 1986). Correspondence, including requests for reprints, may be addressed either to James Crouse, Department of Educational Studies, Willard Hall Education Building, Room 219, University of Delaware, Newark, DE 19716, or to Linda S. Gottfredson, Center for Social Organization of Schools, Johns Hopkins University, Baltimore, MD 21218.

admissions (1985a, 1986; Jencks & Crouse, 1982). He argued that the Scholastic Aptitude Test (SAT) does not help colleges make better admissions decisions, contrary to claims by the Educational Testing Service (ETS), which publishes the test. He recommended that colleges now requiring the SAT abandon that test and switch to standardized achievement tests for selecting freshmen.

Crouse's argument is of particular interest because it is not an attack on the quality or fairness of standardized tests in general or of the SAT in particular, as most previous criticisms of mental tests have been. Not only does he suggest replacing the SAT with another type of mental test, but Crouse himself has also provided evidence for the construct and predictive validity of both aptitude and achievement tests. Furthermore, he has amassed more evidence to support his conclusions than have others who have also recommended dropping the SAT or substituting achievement tests in college admissions (e.g., Slack & Porter, 1980).

Crouse's shift of concern from the *validity* to the practical *utility* of tests is part of a little appreciated but growing trend in scientific debates over testing. The same trend has appeared for mental tests as for other types of assessment during the last decade (see Gottfredson, 1986b, on vocational interest assessments; see Wigdor & Garner, 1982, and Hunter & Schmidt, 1983, on ability tests for education and personnel selection). Thus, his arguments may have significance for testing in general, not just for the SAT. The more general question that Crouse poses, implicitly, is, even if a test is unbiased and predicts desired criteria well, how should it be used in practice, if at all?

The first section of this article reviews research and arguments that demonstrate that Crouse's concern is not principally with the validity or fairness of aptitude and achievement tests. His research provides additional support for the conclusions reported elsewhere in this journal issue that all such tests measure the same *g*, and that it is mainly their success in measuring that *g* that accounts for their value in predicting occupational performance and success. The second section summarizes Crouse's argument that colleges should stop requiring the SAT for admission and consider using achievement tests instead. The third section organizes under several rubrics the major debates about Crouse's recommendations that require additional research for their resolution. Counterarguments were culled primarily from publications by Crouse (1985a, 1985b; Jencks & Crouse, 1982), from defenses of the SAT and admissions testing (e.g., Educational Testing Service, 1980; Hanford, 1985; Jackson, 1980; Klitgaard, 1985; Linn, 1982; Manning & Jackson, 1984) and from relevant debates between participants during the Southern California Personnel Testing Conference. The article concludes by examining the role of science in decisions concerning test use.

CROUSE ON THE PREDICTIVE VALIDITY OF APTITUDE AND ACHIEVEMENT TESTS

During the last decade Crouse has used various longitudinal studies of students and adults to examine the validity of aptitude and achievement tests for predicting several types of adult success, including years of education attained, occupational status, and earnings (Crouse, 1979; Olneck & Crouse, 1979). Countless studies have looked at the relation of "ability" to education, status, or income, but Crouse's research has been unique in the number of data sets examined and in the systematic attention given to the predictive validity of alternative types of mental tests. His results are consistent with the large body of research on attainment. For example, he shows that academic ability is highly correlated with years of education attained, moderately so with occupational status, and less so with earnings. One special contribution of his research, however, has been to examine the impact of using alternative definitions and measurements of academic ability.

Tests of academic ability are generally classified as either aptitude or achievement tests. Aptitude tests (e.g., tests of verbal aptitude) are not tied explicitly to any particular academic curriculum and are made up typically of a wide variety of item types. In contrast, achievement tests (e.g., tests of knowledge in biology) target a narrower band of item content and are tied to particular curricular areas.

In one major study, Crouse factor analyzed 30 cognitive tests in the Project Talent battery. Via correlation and regression analyses he attempted to clarify the relations among (a) the individual tests, (b) the major factors emerging from the tests, and (c) criteria of educational, occupation, and economic success. Subsets of tests were created randomly as well as by grouping tests according to whether they are viewed primarily as tests of rote memory, ability/aptitude, knowledge of specific academic subjects, or knowledge of nonacademic subjects. Academic achievement tests are represented by the third type of test.

Crouse (1979, p. 91) concluded that his findings "suggest that academic ability is largely but not entirely one-dimensional, at least for predicting life chances. 'Non-academic' tests seem to predict success only insofar as they correlate with academic tests." His research finds the same general academic ability factor from factor analyses of all subsets no matter how the subsets of tests are created, as long as there are 10 or so of them. Moreover, these first principal components, or *g* factors, are essentially the same whether the samples used are mostly black or white or rich or poor. Both the aptitude and achievement tests are reasonably good measures of *g*, as is usually the case unless educational opportunities are limited within a population, say, as in a developing country. These conclusions, therefore, are entirely consistent with those

of other participants in the conference (see Jensen, 1986; Hunter, 1986; Thorndike, 1986; Gottfredson, 1986a).

The tests covering a wide range of academic abilities predicted later educational and economic success better than did tests covering a single ability. This finding is consistent with research by Hunter (1986) on the dominance of general over specific abilities in the prediction of performance on the job and in training. Furthermore, the most highly *g*-loaded tests were the best predictors of later success; they were also the tests on which the standardized black–white mean differences were the largest. These results are consistent with the recently revived Spearman hypothesis, which predicts substantial correlations between the *g* loadings of tests and the size of the black–white mean differences that they manifest (see Jensen, 1985, 1986). Finally, for our purposes here, Crouse (1979, p. 85) concluded that “tests given as early as sixth grade appear to predict educational attainment, occupational status, and earnings as well as tests given later.” He goes on to suggest that “it is not cognitive skill per se that affects later success. Rather, the stable motivations and aptitudes that lead to the development of cognitive skills also affect later success. A test’s predictive power appears to derive in large part from its relationship to these stable underlying factors.” This conclusion provides additional support for criticisms by Jensen (1984) of the “specificity doctrine” in psychometric and personnel selection research, which holds that ability tests measure the accumulation of different bits of knowledge or discrete skills rather than some underlying capacity that, among other things, determines the speed of accumulating such bits of knowledge.

Although Crouse has provided persuasive evidence that aptitude and achievement tests are both highly *g* loaded and that they have considerable predictive validity in important social arenas, it should be noted that Crouse has been agnostic about the causal significance of *g* and the predictive relations he has documented. Like many other scholars concerned with issues of social inequality and social policy, he has raised the possibility that general academic ability need not be linked to success, and that its predictive validity may depend on existing practices among educators and employers that reward academic ability. While important, the question of causality is not at issue here and is brought up only to clarify the point that predictive validity neither implies or requires arguments about causal importance. (The issue is discussed further in Gottfredson, 1986a).

The foregoing research did not include either the SAT or ETS achievement tests involved in college admissions, but there is no reason to believe that the results are not generalizable to them. With regard to the SAT, Crouse has noted the vast amount of evidence showing that it has at least moderate validity for predicting college outcomes. Crouse also acknowledges evidence that SAT scores and high school rank together

provide better prediction of college performance than does rank alone. Jencks and Crouse (1982) found no studies in 1982 comparing the predictive validity of the SAT versus achievement tests, but they marshalled various kinds of indirect evidence to support their contention that the SAT is not superior to achievement tests for predicting college performance.

CROUSE'S POSITION ON THE SAT IN COLLEGE ADMISSIONS

Some of the things Crouse is *not* contesting should be made clear at the outset. He has not argued against the principle of selection itself. Some people do. He has not argued for either a greater or lesser emphasis on academic ability in the college admissions process; instead, he implicitly accepts the legitimacy of using indicators of academic competence in some capacity for making admissions decisions. In short, he examines the utility of the SAT in the context of prevailing values and beliefs about the admissions process, which includes the belief that academic ability should be neither completely ignored nor the exclusive basis of admissions decisions.

Crouse's Argument for the SAT's Redundancy in College Admissions

Although ETS promotes the SAT primarily because its use purportedly improves the quality of college admissions, Crouse estimates that the practical benefits of the SAT are minimal for this purpose. The reason is that the SAT, when used efficiently in combination with applicants' high school records, provides predictions of success that are largely redundant with those made from the high school record alone. The test's efficient use as a supplement to the high school record therefore changes few admissions decisions and thus has little effect on admissions outcomes.

Estimating the SAT's benefits. Crouse uses four criteria to assess the quality of the students admitted by a college: the proportion of students with successful freshman grade point averages, the mean freshman grade point average, the proportion of students who finish college and earn a bachelor's degree, and the number of years of college and graduate school that students complete. He compared three hypothetical freshman classes, one "admitted" using a prediction equation based on high school rank alone, one admitted using a prediction equation based on both SAT scores and high school rank, and one admitted randomly. The hypothetical applicants in his major analysis were the 2470 respondents in the National Longitudinal Study (NLS) of the high school class of 1972 who had taken the SAT and who had attended college. Variations of these three policies were also examined, including differences in the minimum predicted grade point averages required for admission (2.5 vs 3.0) and in the selection ratio (0.80 vs 0.20). By estimating the incremental validity provided by using the SAT in addition to rank, Crouse thereby estimates the *maximum* marginal value of the SAT over efficient use of high school records alone.

He notes that the SAT's maximum marginal utility may seldom be realized in practice, however, because optimum prediction strategies are rarely employed by colleges, and, in addition, many of them accept by far most of the students who apply.

No matter which set of assumptions Crouse used in simulating improvements in the academic quality of students admitted, the benefits appeared to be uniformly low. When freshman grades are the success criterion, adding the SAT to the high school record never increases correct admissions decisions by more than 1–3 per hundred or the average grades of the admitted freshmen by more than 0.03 on a 4-point scale (see Crouse, 1985a, Tables 5 and 8). When college completion is the success criterion, the gains in correct admissions decisions are only 1–3 per thousand, and the average increase in years of education is only about 0.01 (see Crouse, 1985a, Tables 5 and 8). Declines in a variety of admissions errors, such as false positives and false negatives, are also similarly small.

The major reason that the benefits are so low, Crouse argues, is that the SAT provides redundant information. He calculates that typically at least 90% of the decisions to admit or reject are the same whether or not the SAT is used in conjunction with high school rank. The reason for this is that SAT scores and high school rank are moderately correlated (.4–.5) with each other and with educational outcomes, so that outcomes predicted from high school rank alone have a part-whole correlation of at least .8 with outcomes predicted from rank plus SAT.

Crouse also used data from other sources to check the generalizability of his conclusions. Analyses with data from the ETS Validity Study Service on freshmen classes admitted to various colleges and universities also suggested that using the SAT plus rank rather than rank alone produced no more than 2 to 3 more correct admissions decisions per hundred applicants.

None of Crouse's research to date should be used to argue that there are no other benefits to colleges' use of the SAT. For example, another argument has been that requiring the SAT produces more highly *self*-selected pools of applicants because it influences applicants' beliefs about their chances of being admitted to different colleges. Such a virtue of the SAT, if true, could lessen the processing burden upon admissions committees. Crouse points out, however, that no evidence has been put forward by ETS to support this suggested benefit of the SAT.

Another advantage the SAT is presumed to have over grades and achievement tests is that performance on the SAT is less susceptible to variations in educational opportunity and school quality, and it therefore allows schools to identify "diamonds in the rough" that would be denied admission on the basis of other information alone. Crouse has questioned whether there really are very many such individuals. Jencks and Crouse

(1982) also imply that the SAT is not insensitive to environmental quality when they argue that achievement test performance is no more responsive to differences in school quality than is performance on the SAT. Their implication is certainly consistent with repeated statements issuing from ETS that the SAT is not a test of innate ability and with ETS's frequent attributions of group differences in SAT scores to failures of educational and social policy (Manning & Jackson, 1984; Educational Testing Service Board of Trustees, 1984; Jackson, 1980; see also Gordon, in press). (No mental test is recognized as a direct measure of innate abilities and there is considerable debate about whether school quality is a significant source of group differences in SAT scores.)

Costs of the SAT. Crouse has made no formal estimates of cost, but he has listed several types of cost: the fees paid by the 1.5 million students every year who take the SAT (some of whom will take it more than once), the fees some people pay for coaching on the SAT, and the anxiety of taking a test that may influence one's life chances. These are costs borne by test takers themselves. But there is also the systemic possibility that requirements to take the SAT for college admission "undermine efforts to improve secondary education" (Jencks & Crouse, 1982, p. 32). The argument is that the SAT constitutes a disincentive for learning in high school. First, it emphasizes basic cognitive skills and "by emphasizing skills that secondary schools do not explicitly teach, at least after tenth grade, the SAT implicitly tells secondary schools that most of what they teach does not really matter" (p. 35). Second, the SAT induces complacency and resignation because the word "aptitude" in its name encourages students to think that there is nothing they can do to increase their performance on the test; therefore, they have no incentive for learning. In short, "because of its content and its name, the SAT does not appear to reward diligence" (p. 35).

Crouse's Argument for Substituting Achievement Tests for the SAT

Crouse's major argument for substituting achievement tests for the SAT is that they appear no worse than the SAT, but they may be better in at least one important way. Specifically, Crouse argues that there is no reason to presume that achievement tests are any less predictive of success, that they have more adverse impact on disadvantaged groups, or that they are any more susceptible to differences in school quality than is the SAT. "In most colleges, scores on ETS achievement tests are almost interchangeable with SAT scores" (Jencks & Crouse, 1982, p. 28).

Benefits. The distinctive benefit of achievement tests, Crouse argues, is that more than the SAT they encourage diligence in high school and they encourage learning beyond the basics. He argues that perhaps one reason European and Japanese teenagers work hard is that at the end

of secondary school they must take achievement tests that largely determine whether and where they will be allowed to obtain higher education. If confronted with achievement tests rather than the SAT, American teenagers might work harder than they do now.

A preeminent justification for requiring the SAT in college admissions has been that it provides an objective yardstick of academic ability and thereby accords with the prevalent belief that people should be selected and rewarded according to merit. Jencks and Crouse (1982) suggest that the use of achievement tests would be more consistent with a competing principle of merit and distributive justice to which Americans also commonly subscribe, namely, that it is more just to reward effort than traits over which people assume they have no control (i.e., "smarts").

Costs. Crouse concedes that one problem with achievement tests, if widely adopted for college admissions, is that they could have a big influence on secondary school curricula and perhaps lead to a de facto national curriculum—outcomes to which many schools, teachers, and parents would heatedly object. However, he argues that whether this would happen depends upon how a system of achievement tests for college admissions is implemented. It is not a logical necessity.

DEBATE OVER CROUSE'S PROPOSALS

The exchange between Crouse (1985a, 1985b) and Hanford (1985), who is the president of the College Board, includes many of the counterarguments to Crouse's position. According to Hanford, (a) Crouse understates the relation between the SAT and college outcomes, (b) his results do not hold for selective colleges, (c) his high correlations between outcomes predicted in different ways are deceptive, (d) he uses narrow criteria for assessing the quality of college admissions (focusing excessively on academic ability), and (e) he oversimplifies the college admissions process.

Crouse rebuts each of these criticisms in his reply to Hanford (Crouse, 1985b). Furthermore, Crouse charges that (a) the College Board ignores the evidence from its own studies that corroborates Crouse's conclusions, (b) the Board fails to warn colleges of the problems of predicting applicants' college performance from equations based on previously admitted classes, (c) the Board ignores the need for evidence to support the benefits it suggests the SAT has (i.e., its use can improve the quality of admissions), and (d) the Board fails to investigate the potential costs and benefits of the SAT. Crouse (1985b, p. 485) adds that his "growing concern is that the College Board and ETS perhaps cannot be trusted fully to appraise the usefulness of the SAT or to reply openly and honestly to criticism." He then suggests that ETS's Validity Study Service could improve evaluation of the SAT as a tool in admissions by helping individual colleges estimate the SAT's marginal benefits in the same way that Crouse has. Colleges could then have improved information, he argues, to evaluate for themselves the utility of the test.

These arguments and counterarguments illustrate current disagreement about how to estimate the magnitude of any increases in quality of admitted students, about the practical significance of small increments in benefit, about what types of costs and benefits should be considered, and perhaps about whose responsibility it is to evaluate the utility of the SAT. In an effort to clarify some of the more fundamental dimensions of the debate over Crouse's recommendations, we have organized the issues into five general categories. In presenting them, the focus has been on aspects of the issues that could be clarified by further research.

1. *The SAT's incremental value in admissions.* The dispute here is not about whether the increments in student outcomes attributed to the SAT have been underestimated, but whether small increments are important in a practical sense. It is always hazardous to dismiss "small" increments as inconsequential simply on the basis of their absolute magnitude. If, for example, two or three additional errors per hundred applicants translates into expensive counseling and academic assistance costs, these small incremental gains could be important. Likewise, if the two or three additional correct admissions decisions per hundred translates into a few extra superior performers who boost a college's academic climate and future reputation, the small incremental gains could again be important.

Crouse doubts that either of these outcomes is likely, but points out that additional research, not opinion, is required to settle this issue. The College Board and ETS could exercise leadership in this regard by assisting individual colleges to estimate the costs of small increases in admissions errors in a currency other than freshman grades and graduation rate. Colleges would then be in a better position than they are now to estimate the practical utility of the SAT.

2. *Simulation of the admissions process.* The issue here is whether the model of admissions used by Crouse leads to an understatement of the magnitude of the criterion changes (as distinct from the practical value of any given amount of criterion change). All methodological questions raised about Crouse's treatment of the data fall under this heading. So do criticisms that Crouse's simulations do not adequately represent benefits of the SAT for highly selective institutions. Crouse has discussed these criticisms in his reply to Hanford (Crouse, 1985b). These issues all deserve further study, but we use this space to draw attention to an additional issue that badly needs investigation.

Standardized tests of academic ability may prevent admissions committees from straying too far from a concern with academic ability. For example, students are often selected in an essentially judgmental or clinical manner with few or no explicit guidelines. Furthermore, admissions committees sometimes pay considerable attention to nonacademic traits and accomplishments, especially when selecting undergraduates, that presumably contribute not only to well-rounded individuals but also to

a beneficial diversity within the student body. One is led to wonder, then, how often grades and other evidence of academic accomplishment in high school are used in an optimum manner. Conceivably, the mere presence of standardized test scores directs relatively more attention to academic ability than would occur otherwise. Without standardized test scores, the admissions process in many colleges might veer even more toward a mystical weighing of the wide variety of nonacademic attributes that admissions committees often value but that have no demonstrated relation to success in college. The point here is that the presence of SAT scores may promote more optimum use of grades in the admissions process. This hypothesis is a variant of the common argument that the SAT is an integral part of a system of checks and balances in the evaluation of applicants (Hanford, 1985). If this is true, the question then becomes one of whether aptitude tests such as the SAT serve this function better than would achievement tests. No one has explored this question.

3. *Other benefits to colleges of the SAT.* SAT scores often help colleges and others to monitor the quality of educational programs and the workers entering different professions. Mean SAT scores within an institution or professional program are probably the most common single indicator of the quality of those programs. Although this measure of quality best refers to the general competence of the people receiving instruction rather than to the instruction they receive (despite statements to the contrary by ETS and others), it does constitute one measure of overall quality for many practical purposes (such as for attracting prospective instructors or employers of students). Unfounded prejudice seems insufficient to explain the watchful eye that administrators keep on the SAT scores of their students relative to those in other programs or in previous years, or to explain the great concern that is generated by falling SAT scores among students entering particular professions.

Colleges do, of course, need some ways to monitor their programs. Crouse argues, however, that the hegemony of the SAT for this purpose has distracted attention from the question of how colleges *should* monitor program quality. He fears that the SAT has been used for this purpose primarily because it is an easy and available indicator, and he asks, for example, whether nationally normed achievement tests might not allow administrators to monitor and diagnose problems as well as or better than the SAT. Achievement tests have the virtue, he continues, of being tied to the specific curriculum content that must be monitored or diagnosed. Consequently, if nationally normed achievement scores of students entering a particular program rise or fall, these changes are likely to be more informative than any changes in the more limited range of verbal and mathematical skills measured by the SAT. This issue also deserves further study.

4. *Additional benefits for college applicants.* ETS and the College Board argue that the SAT can help identify "hidden" talent or "diamonds

in the rough”—able persons who have performed poorly in school but who might do well in college if given the chance. However, Jencks and Crouse (1982) have argued that the SAT seems no less responsive than achievement tests to environmental differences, because the between-school variance in performance is no greater for achievement tests than it is for the SAT. Crouse also argues that even if there were merit to the “diamonds in the rough” argument, the number of such persons is unknown and has never been shown great enough to justify the widespread testing with the SAT.

Another possible benefit is that the SAT may facilitate decision making by applicants and their parents, because it provides a yardstick to assess the wisdom of applying to and accepting offers from alternative schools. The debate is about whether, given everything high school students already know about themselves and different colleges, the additional knowledge of SAT scores really influences their perceptions about where they will be most successful. For example, if applicants expect to do well in a specific college because they received high grades in high school, are these expectations for success lowered by learning that their SAT scores are low relative to those of other students typically admitted to that college? Are students with good high school records but low self-confidence reassured by high SAT scores and thereby more likely to pursue more demanding educational options? Are less wealthy parents more willing to invest in more expensive options when they have additional evidence concerning their children’s chances of success in such settings? Providing reassurance and facilitating wiser choices are commonly viewed as benefits of vocational interest inventories, but it is hard to know how seriously to take these as potential benefits of the SAT in the absence of relevant evidence.

5. *Testing for selection vs performance appraisal.* Crouse has argued that substituting achievement tests for the SAT in college admissions may help create desirable incentives for high school students and have a beneficial effect on secondary school curricula. He argues that both the SAT and achievement tests may function most usefully, not as highly beneficial decision-making tools for colleges and applicants, but as interventions in high schools, and that the impact of achievement tests might be surprisingly positive. The question that arises is whether the same tests and assessments can be used effectively and simultaneously for the evaluation and motivation of high school performance and for college admissions, that is, for performance appraisal as well as for selection.

Naturally, there is little disagreement that improving students’ high school performance is desirable, that students would profit from working harder, or that achievement tests would have a substantial impact on high school curricula. Indeed, performance assessment is generally viewed

as a potentially effective tool for directing behavior in some directions rather than others, both in schools and on the job. However, because widespread achievement testing for college admissions has been limited to the American College Testing Program, which uses only four tests, a great deal of uncertainty surrounds the possible consequences of colleges shifting to a larger number of achievement tests for admissions purposes.

One potential problem is that some standard set of achievement tests seems desirable for selection because of the economy and comparability it provides applicants and colleges in the admissions process. However, such uniformity in assessment may be seen widely as unsuited to the diversity of goals and circumstances typifying the many schools in this large nation unless a very large number of tests is used. Any set of tests is therefore likely to bring about charges of infringement on some traditional local and state prerogatives to determine primary and secondary school curricula. Local schools complain now that state-imposed tests such as the New York State Regents Examinations distort instruction in the later high school years.

Whether a nationwide admissions program based on achievement tests would be politically acceptable thus remains to be seen. Crouse's own view is that observers typically overstate curricular differences among high schools' preparation for college academic subjects, and that achievement tests fair to students educated under differing curricula can be prepared. ETS now claims, for example, that its existing achievement tests are fair in this regard. If this is so, then there would seem to be no reason why they could not continue to be fair.

Another potential problem is that if achievement tests required for college admissions were to become viewed as an extension of the secondary school curriculum, the number of stakeholders in the results on those tests would multiply drastically. The content and difficulty level of the tests could become subject to considerable negotiation among diverse and competing constituencies. The number of different tests could be expected eventually to multiply in order to accommodate desires for representing fairly the diversity of local goals and values. Difficulty levels and discriminability of the tests would probably shift in response to changing curricular goals, to debates about the appropriateness of criterion-referenced versus norm-referenced tests, to debates about the relative importance of nonacademic skills or nontraditional academic subjects (e.g., black history), and to differences in score distributions among racial/ethnic groups. In short, testing for college admissions would become more highly politicized than ever and perhaps less valid. Confusion caused by proliferating tests might also eventually undermine the value of achievement tests for making better or more expeditious college admissions decisions.

Crouse argues that these negative consequences may be overstated. For one thing, he continues, achievement tests ought to evolve constantly in response to changing curricular philosophies and in response to continual political and social renegotiation of who and what should be rewarded. The College Board has tried to serve as a forum from its outset to influence this evolution toward a tighter articulation between college preparation in high school and what colleges want their entering freshmen to know and be able to do. Furthermore, the specific, distinct, and trainable skills and knowledges taught in high school for college preparation ought to be debated. They are debated now. The main thing that measuring these skills would do is create additional incentives to learn them. If the evolution of achievement test content were to improve the match between the skills high school students are taught and the skills they need in college, then there is no reason to believe that the predictive validity of achievement tests for admissions purposes would be impaired. Given that the SAT does not now do what it is intended to do, and given the feasibility of basing college admissions testing on high school students' mastery of the high school curriculum, such a system of achievement testing seems worth exploring further.

CONCLUSION

Public controversy has surrounded mental tests from early in this century (Haney, 1981). The last two decades have seen a barrage of challenges concerning their validity, with special scrutiny being devoted to the issue of bias against different racial-ethnic groups. As extensive scientific evidence has accumulated concerning their validity, controversy over testing seems to have shifted more toward questions of utility.

Crouse's questioning of the ETS claim that the SAT leads to better college admissions decisions is a good example of this new direction in criticism. He grants more to the SAT than most of its critics have, specifically, that the test has substantial validity for predicting educational and later life outcomes. At the same time, however, he asks whether the SAT is really *useful* in applied settings. Does it add anything to the admissions process except additional cost and anxiety for applicants? Might its use actually be detrimental to secondary education in the United States? If so, what are the alternatives? If science outdistances polemics in the debate over the foregoing questions, clarification of important and ever troublesome testing and selection issues may result.

An examination of the costs and benefits of using different standardized tests in college admissions, or of not using them at all, could profitably involve a look at the experiences of educational systems that have pursued alternative selection strategies or that have changed them over time. For example, even though they are of admittedly limited generalizability, what have been the experiences of American colleges that have adopted

achievement tests in preference to the SAT? Also, why have fewer colleges opted in the past for achievement tests than for the SAT? Japan, France, and Great Britain are often cited as countries that use achievement tests, for better or worse, but Australia might provide a particularly interesting set of case studies.

In Australia at present (mid 1986) there are no private universities and, apart from the Australian National University in Canberra, all the state colleges and universities draw their students primarily from the secondary schools situated in the same state. Examination and college admissions practices are uniform within states, generally being set by state boards of education, but the practices differ widely across states. At present, for example, one state evaluates college applicants largely according to results on uniform statewide achievement tests; another uses secondary school grades and achievement test results with approximately equal weight; three others use grades and the Australian SAT. Also, some states have changed their practices over time in response to problems arising from the earlier practices.

Progress in assessing the utility of tests also would be facilitated by more analytical probing of the consequences of their various suggested uses. Four questions emerge from the argument and counterargument reviewed in this article. First, what is the main *purpose* for using the test in question—selection or performance appraisal? Under performance appraisal, more specific aims would include the diagnosis of individual needs, the motivation or redirection of behavior, and program evaluation. Second, in what sort of selection or appraisal *system* is the test embedded? For example, what other tools (e.g., high school record) are currently or potentially available for helping to meet the same goals? And how stable are the properties of these other tools with and without the presence of the test? Third, who are the major *stakeholders* (e.g., individuals, colleges, secondary schools, professional groups, employers, minority groups, majority groups), and how might they react in both the short and long run to particular changes in test design and use? Fourth, what *criteria* (costs and benefits) should be considered in assessing overall utility? Criteria should be specific and measurable where possible. Decisions about test use rest on more than the answers to these four questions, of course, because choices depend on the values one attaches to the different potential costs and benefits. But these questions provide a broad, systemic perspective that can promote greater awareness of the ultimate consequences of different choices and thus contribute to better informed decisions.

It has become commonplace to note that decisions about how to use mental tests, if at all, reflect and influence the nature of the society we live in. Haney's (1981, p. 1032) closing remark is a typical one in professional discussions of the uses of mental tests: "while the social role of standardized

testing is both advocated and challenged in technical terms, the prominent social concerns surrounding standardized testing, both now and in the past, are rooted in matters of social and political value." The inference that one might mistakenly draw from such statements is that the proper province of scientists does not extend beyond the scrutiny of the psychometric properties and appropriate interpretation of mental tests to encompass critical examination of choices involving values. By default, that leaves examinations of such choices to persons less equipped by training and experience with testing to conduct them. Although it may not be the proper role of social scientists to negotiate the social and political values involved in testing, it certainly does fall within their special competence and hence responsibility to clarify past or probable consequences of using tests in particular ways for particular purposes.

REFERENCES

- Crouse, J. (1979). The effects of academic ability. In C. Jencks, S. Bartlett, M. Corcoran, J. Crouse, D. Eaglesfield, G. Jackson, K. McClelland, P. Mueser, M. Olneck, J. Schwartz, J. Ward, & J. Williams (Eds.), *Who gets ahead? The determinants of economic success in America* (pp. 85–121). New York: Basic Books.
- Crouse, J. (1985a). Does the SAT help colleges make better selection decisions? *Harvard Educational Review*, *55*, 195–219.
- Crouse, J. (1985b). This time the College Board is wrong. *Harvard Educational Review*, *55*, 478–486.
- Crouse, J. (1986). Should a million and a half students be required to take the SAT next year? *Phi Delta Kappan*, *67*, 346–352.
- Crouse, J., & Trusheim, D. (1986). *The case against the SAT*. Manuscript in preparation.
- Educational Testing Service (1980). *Test use and validity: A response to charges in the Nader/Nairn report on ETS*. Princeton, NJ: Author.
- Educational Testing Service Board of Trustees (1984). *Educational Testing Service Trustees' 1984 public accountability report*. Princeton, NJ: Educational Testing Service.
- Gordon, R. A. (in press). Jensen's contributions concerning test bias: A contextual view. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy (Falmer International Master-Minds Challenged)*. Sussex, England: Falmer Press.
- Gottfredson, L. S. (1986a). Societal consequences of the g factor in employment. *Journal of Vocational Behavior*, *29*, 379–410.
- Gottfredson, L. S. (1986b). Special groups and the beneficial use of vocational interest inventories. In W. B. Walsh & S. Osipow (Eds.), *Advances in vocational psychology: Vol. 1. The assessment of interests* (pp. 127–198). Hillsdale, NJ: Erlbaum.
- Haney, W. (1981). Validity, vaudeville, and values: A short history of social concerns over standardized testing. *American Psychologist*, *36*, 1021–1034.
- Hanford, G. H. (1985). Yes, the SAT does help colleges. *Harvard Educational Review*, *55*, 324–331.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*, 340–362.
- Hunter, J. E., & Schmidt, F. L. (1983). Quantifying the effects of psychological interventions on employee job performance and work-force productivity. *American Psychologist*, *38*, 473–477.
- Jackson, R. (1980). The Scholastic Aptitude Test: A response to Slack and Porter's "critical appraisal." *Harvard Educational Review*, *50*, 382–391.

- Jencks, C., & Crouse, J. (1982). Aptitude vs. achievement: Should we replace the SAT? *The Public Interest*, *67*, 21–35.
- Jensen, A. R. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures*, *7*, 93–118.
- Jensen, A. R. (1985). The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, *8*, 193–219.
- Jensen, A. R. (1986). *g*: Artifact or reality? *Journal of Vocational Behavior*, *29*(3), 301–331.
- Klitgaard, R. (1985). *Choosing elites: Selecting the "best and brightest" at top universities and elsewhere*. New York: Basic Books.
- Linn, R. L. (1982). Admissions testing on trial. *American Psychologist*, *37*, 279–291.
- Manning, W. H., & Jackson, R. (1984). College entrance examinations: Objective selection or gatekeeping for the economically privileged. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 189–220). New York: Plenum.
- Olneck, M. R., & Crouse, J. (1979). The IQ meritocracy reconsidered: Cognitive skill and adult success in the United States. *American Journal of Education*, *88*, 1–31.
- Slack, W. V., & Porter, D. (1980). The Scholastic Aptitude Test: A critical appraisal. *Harvard Educational Review*, *50*, 154–175.
- Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior*, *29*, 332–339.
- Wigdor, A. K., & Garner, W. R. (Eds.). (1982). *Ability testing: Uses, consequences, and controversies: Pt. I. Report of the committee*. Washington, DC: National Academy Press.

Received: July 1, 1986.