



# Trouble Over Testing

Walt Haney

---

*Current conflicts over test bias and "truth in testing" reflect differing aims of education and the increasingly important social role of testing in our society.*

---

Standardized testing is much in the news. Even more than earlier debates—in the early 1960s, concerning personality tests, and in the 1970s over bias in tests—the current furor over standardized testing has become both highly public and intensely political. Testing is being debated in state legislatures and in the U.S. Congress. It is being covered widely in the popular press and on TV and radio. Federal courts more and more are asked to adjudicate testing controversies. The popular prominence of current debate over testing was perhaps epitomized by the appearance of Ralph Nader on the "Tonight Show" on January 24, 1980, to publicize the recent Nader report on the Educational Testing Service (ETS). After condemning the "reign" of ETS, Nader gave an impassioned plea for wider consideration of traits like perseverance, wisdom, idealism, and creativity—traits that cannot be measured by multiple-choice aptitude and achievement tests of the sort ETS publishes. The "Tonight Show" audience broke into spontaneous applause.

The incident symbolized the depth of feelings over standardized testing, but it barely scratched the surface of the myriad concerns and conflicts currently swirling around testing. Should students have to pass minimum competency tests in order to pass each grade or graduate from high school? Should "truth in testing" legislation be passed to guarantee that test-takers have access to corrected test results and to force test sponsors and developers to be more accountable for the instruments they produce and profit from? Should the role of standardized tests as gatekeepers to educational and occupational opportunities be enhanced or diminished? And whatever their role, how can tests be made more fair, to enhance equality of opportunities for both poor and rich, male and female, black and white, young and old? I will explore these questions by comparing and contrasting some of the recent troubles over testing.

Dozens of states around the country have implemented competency testing programs so as to provide a new standard for award of the high school diploma, to guide grade-to-grade promotion, or to identify students for remedial instruction. Though rationalized in several different ways, minimum competency testing programs seem in large measure to represent a worry that all is not well with America's schools, but they may be as much a symptom of the worry as a solution to it. In many places competency testing programs are being implemented without enough attention to how well the tests match the goals and curriculum of schooling and remarkably little consideration has been given to how instruction should be organized to help students who do poorly on the tests. These and other controversies concerning minimum competency testing were addressed most clearly in the federal

court case of *Debra P. v. Turlington* in Florida—a case brought before the court on behalf of ten black students and others who were to be denied diplomas after failing Florida's functional literacy test. Data presented during the trial indicated that had the diploma sanction been enforced, 20 percent of black high school seniors in Florida would have been denied diplomas as compared with only 2 percent of white high school seniors. After a trial of several weeks, the federal judge in the case ruled in July 1979 that the Florida testing program was unconstitutional because it perpetuated the effects of past discrimination and because it was implemented without an adequate phase-in period. As a result, he ordered a moratorium until 1982 on use of the Florida test to deny students a regular high school diploma. The Florida case is now under appeal, but no matter how it is resolved, minimum competency testing at the elementary and secondary level will probably provoke more controversy over the social role of testing.

### "Truth in Testing"

The "truth in testing" movement is of more recent vintage than the competency testing movement, but it is generating no less debate. As with competency testing, "truth in testing" is an ill-defined term. It refers loosely to a variety of efforts to regulate standardized testing, many of which have taken the form of legislative proposals to require that:

(1) Individual test takers have access to corrected test results within a specified period after test administration; (2) Test sponsors and/or publishers file information on test development, validity, reliability, and cost with government agencies; and (3) Testing agencies give individual test takers information on the nature and intended use of tests prior to testing, and guarantee their right of privacy concerning their own test scores. The common rationale for such proposals is threefold. First, advocates argue that as a matter of simple justice, test-takers should have the opportunity to review the corrected results of tests which have important consequences for their educational and job opportunities—they should know the basis on which they are judged. Second, proponents of truth in testing maintain that test sponsors and publishers should be more accountable for the intellectual quality and social consequences of the tests they produce, and that public disclosure will help bring that about. Third, they argue that the limitations of standardized tests—the fairly wide margin of error implicit in any individual score, the limited validity of test scores to predict things like future educational attainment and job performance, and the fact that test scores typically correlate with measures of social and economic background—should be more widely known, especially to individuals who take the tests.

When one compares the "truth in testing" movement with that for competency testing, two sharp differences become apparent. First, while the new



Photo: Michael J. Sexton

---

*"Should the role of standardized tests as gatekeepers to educational and occupational opportunities be enhanced or diminished?"*

---

minimum competency testing programs have focused almost exclusively on *elementary* and *secondary* education, almost all the "truth in testing" legislative proposals have focused on *postsecondary* admissions testing.<sup>1</sup> Second, while the competency movement has clearly *extended* the public policy reach of standardized testing, the "truth in testing" movement has sought to *curtail* or at least to control it.

A similarity between the two is that the "truth in testing" movement also has led to litigation. Most of the controversy surrounding the New York law, which went into effect on January 1, 1980, has centered on the requirement that questions and answers determining examinees' scores on postsecondary and professional school admissions tests be disclosed within thirty days of the release of test scores. Disclosure of test contents might seem fairly innocuous to the uninitiated. Just as teachers usually return corrected papers to their students, it would seem reasonable to ask test sponsors to return corrected test results to individual test-takers. But some publishers of secure tests view the disclosure requirement as a threat to the very existence of systematic admissions testing. As a result of such concern, the Association of American Medical Colleges (AAMC) brought suit against the state of New York on the grounds that there are a limited number of high quality questions that can be asked on a test like the Medical College Admissions Test (MCAT) and charged that the New York law violated its rights protected under federal copyright law. On January 21, a federal judge ordered a preliminary injunction exempting the MCAT from the New York law until the legal merits of the AAMC case can be tested in court.<sup>2</sup>

### The Reign of ETS

Around the same time, considerable fuel was added to the "truth in testing" fires by the release of a 554-page Nader report entitled *The Reign of ETS: The Corporation That Makes Up Minds*. The report, authored by Allan Nairn, levels numerous criticisms against ETS and the standardized multiple-choice tests it has developed. Among the charges are that:

(1) ETS has been more concerned with organizational self-interest than in the public interest it ostensibly serves as a nonprofit educational organization; (2) Tests like the Scholastic Aptitude Test (SAT) and the Law School Admissions Test (LSAT) have very little power to predict how well students will do in school and even less in predicting anything about life after schooling; (3) That such tests are biased against minority group and lower income students; and (4) That despite the disclaimers of ETS and test sponsors, coaching can be effective in boosting scores on such aptitude and admissions tests. Only hours after a Nader press conference to release the Nairn report, ETS held its own press conference to rebut Nader's charges. ETS President William Turnbull asserted that much of the material in the Nader report is dated and "some of it is just wrong in its conclusions." Turnbull also charged that "Nader and Nairn wrongly blame the tests for showing that minority students are less well prepared in school than majority students. . . . The tests do not create the difference; they reveal it."<sup>3</sup>

Turnbull's assertion is one of the root issues behind all of the current controversies over testing. Do standardized tests present us with a valid and unbiased picture of our social world, or are they like fun house mirrors, reflecting a distorted image of both our world and ourselves?

### The Florida Competency Testing Case

In the Florida competency testing case, these issues were hotly disputed. In discussing test validity, expert witnesses for both the plaintiffs and the defendants frequently referred in their testimony to the *Standards for Educational and Psychological Tests* published by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education in 1974.<sup>4</sup> George Madaus of Boston College, as an expert witness for the plaintiffs, testified that adequate validation of an instrument like the Florida functional literacy test requires the integration of construct validity studies, criterion validity studies, and content

<sup>1</sup> There are, it should be noted, some exceptions to this general pattern. In a few places, minimum competency testing is being extended upwards into postsecondary institutions, and at least one state—Massachusetts—is considering proposed truth in testing legislation which would apply to standardized tests given at the elementary and secondary levels.

<sup>2</sup> "Medical Colleges Win Injunction Against Testing Law," *The New York Times*, 22 January 1980. To avoid clutter, subsequent references to sources listed at the end of this article are omitted. Where relevant, page references to these sources are provided.

<sup>3</sup> "Nader Accuses ETS of 'Fraud,' Plans Drive for Testing Reform," *Chronicle of Higher Education* (January 1980).

<sup>4</sup> *Standards for Educational and Psychological Tests* (Washington, D.C.: American Psychological Association, 1974).

validity studies as described in the *Standards* (1732).<sup>5</sup> Madaus charged that the Florida test had not been properly validated in any of these three areas. Even though Florida officials had contracted to have ETS prepare test item specifications to describe the items on their test, Madaus testified that this did not constitute adequate content validation because both state officials and test developers (the Florida Department of Education, ETS, and National Evaluation Systems, which was also contracted to help develop the Florida tests) had completely failed to consider whether either the test items or the item specifications matched the curriculum covered in Florida schools. When asked what he thought of the use of the Florida functional literacy examination to determine whether or not students ought to receive high school diplomas, Madaus answered simply, "It's unconscionable" (1978).



Photo: Library of Congress

The state of Florida called on its own expert witnesses in educational testing. One witness, who testified at length on the test *Standards* was William Mehrens of Michigan State University. Mehrens, in sharp contrast to Madaus, testified that neither construct validity nor criterion-related validity were necessary for a test such as the Florida functional literacy test (2225, 2227, 2300). Moreover, when asked whether the content validity of the Florida test met acceptable professional standards, Mehrens answered, "It would be my judgment that it does, yes" (2229). Asked whether test content should cover what students were taught in school, Mehrens answered:

... if one believed that what you ought to have is a test that tests those things that have already been taught in school, then that consideration ought to take place prior to the building of the test, that it ought to take place at

the time one is determining what the objectives are that ought to be measured. To build a valid test, a test constructor really has no legitimate option except to build questions on the objectives (2262-63).

On the question of whether the Florida test was biased, the court heard similarly conflicting expert testimony. Lawyers for the plaintiffs called on several experts who identified specific test questions they considered biased against poor and minority students. But lawyers for the state of Florida called on their own experts to defend the test. One of them, James Popham of the University of California, addressed the test bias issue as follows:

The test, the minimum competency program in Florida, is not biased against minority youngsters, against youngsters from low socioeconomic backgrounds. It's just the opposite. It allows one to isolate and thereafter remediate instructional deficits that are currently operative in the state (3190).

Although the judge in the Florida case, George Carr, enjoined the state from forcing students to pass the test as a requirement for high school graduation for a period of four years, he seemed to side with the state of Florida on the specific issues of test validity and bias. Judge Carr held that the Florida test "had adequate content and construct validity and bore rational relation to valid state interest" and that the "plaintiffs failed to establish that the test was racially or ethnically biased" (474 *Federal Supplement*, p. 244).

### The California IQ Testing Case

Just three months later, another federal court judge at the opposite end of the country reached an opposite conclusion about a different kind of standardized test. In the case of *Larry P. v. Riles*, in California, plaintiffs had challenged the placement of students into classes for the "educable mentally retarded" (EMR), and particularly the use of standardized individual intelligence tests in such placement. They contended that in their current form such IQ tests are biased against black children and hence lead to misplacement of black children in EMR classes. Pointing out that black children represent only 10 percent of the general student population in California but provide some 25 percent of EMR enrollments, they contended that misplacement doomed children to "stigma, inadequate education, and failure to develop skills necessary to productive success in society."

The California case lasted far longer than the Florida case. The initial complaint was filed in 1971,

<sup>5</sup> This and subsequent references to testimony in the Florida trial refer to page numbers in the official trial transcript.

and the trial transcript ran to more than 10,000 pages. However, as in the Debra P. case, there was conflicting testimony from expert witnesses; "renowned experts disagreeing sharply" as the court put it. In this instance, however, on the question of test validity and bias, federal court judge Robert Peckham ruled in favor of the plaintiffs, against the state. In an opinion issued October 16, 1979, the judge ruled:

In violation of Title VI of the Civil Rights Act of 1964, the Rehabilitation Act of 1973, and the Education for All Handicapped Children Act of 1975, defendants have utilized standardized intelligence tests that are racially and culturally biased, have a discriminatory impact against black children, and have not been validated for the purpose of essentially permanent placement of blacks into educationally dead-end, isolated, and stigmatizing classes for the so-called educable mentally retarded (p. 3).

So in just the past ten months, two federal court judges have reached quite different conclusions regarding the validity of standardized tests and the

question of whether they are biased against minority students—though both these decisions are currently under appeal. Since both federal court judges and testing experts disagree sharply on these issues, it is not surprising that public opinion is divided as well.

#### Testing: Pro and Con

Two recently published and much publicized books present opposite sides in the public debate on standardized testing. Nader and Nairn's *The Reign of ETS*, mentioned earlier, is a case for the prosecution. Arthur Jensen's new book, *Bias in Mental Testing*, is very much a brief for the defense. The two volumes represent not just different views of standardized testing but very different genres and world views as well.

The Nader report is an attack on ETS, the "General Motors" of the testing industry, as much as it is an attack on standardized testing. The genre is muck-raking journalism, and the report is based on many of

### Update on Test Disclosure Legislation Chris Pipho

Legislation requiring test disclosure is already on the books in two states, has been proposed in 13 others, and is expected in at least four more this year.

California became the first state to require disclosure of commercial test information with a law passed in 1978 that applies to any standardized test given for postsecondary admissions to more than 3,000 students. New York followed in 1979 with the now famous LaValle Bill that requires release to students of the complete test, the scores, and the answers within 30 days after the test. Legislation was introduced in eight other states in 1979, but none of the bills was enacted. In some instances they became the subject of interim study, carry-over legislation, and reintroduced bills.

When this report was assembled, the issue was alive in Connecticut, Indiana, Maryland, Massachusetts, Minnesota, Mississippi, Missouri, New Jersey, Ohio, Pennsylvania, South Carolina, and Tennessee. In all these states the primary thrust of the legislation has been to regulate admissions and other commercial tests used at the postsecondary level, but bills in New Jersey, Massachusetts, and Maryland also apply to testing in elementary and secondary schools. One Massachusetts bill calls for a special commission

to study the College Entrance Examination Board (CEEB) and the use of postsecondary entrance tests in general.

It is difficult to assess the probability of these bills being enacted and even to know the current status of some of them. One state, for example, was rumored to have as many as seven bills in the hopper, but at the time I checked, only three were visible.

Generally the bills are assigned to the appropriate education committee. Some have been heard one or more times, while others will probably die in committee as happened in Oklahoma. Some may become the subject of interim studies.

On the federal scene, the issue has apparently cooled after the Weiss and Gibbons Bills were debated in the fall of 1979. The issue is potentially still alive but no new federal bills have been introduced in 1980.

*Chris Pipho is Associate Director, Research and Information Department, Education Commission of the States, Denver, Colorado.*

the tools of the muckraker's trade: tenacious investigation, leaked internal documents, and hints of guilt by association. Nairn's tenacity is apparent from the simple facts that he has been working on his investigation of ETS for the past six years and that he has interviewed 200 people at ETS, ranging all the way from the ETS President and officers to clerical workers and telephone operators. Some of the passages he quotes from internal memos of ETS and affiliated organizations are wonderfully revealing; for example, a program of "defensive research" regarding the LSAT (p. 254); and an Orwellian suggestion by an upwardly mobile ETS Vice-President in 1963 for a "campaign for verbal purification" to help prevent ETS from being viewed as a crass commercial enterprise (p. 347).

The strongest theme of guilt by association is the connection Nairn recounts between current practices of standardized intelligence and aptitude testing and the social policy recommendations which early mental testers ostensibly derived from the "scientific" study of intelligence. Nairn describes the leadership of several pioneers of mental testing—Galton, Terman, and Brigham, the father of the SAT—in the eugenics movement, which aimed at human improvement through genetic control; specifically via limitations on the reproduction and immigration of "inferior" races. The link, Nairn suggests, was organizational as well as intellectual. "In 1916, the National Council on Education and the National Education Association, two groups which had earlier launched 'the movement which resulted in the organization of the College Entrance Examination Board,' established the 'Committee for Racial Well-Being,' a eugenics strategy group which included testing pioneers Henry Goddard and Robert M. Yerkes" (p. 172).

Revealing though internal documents and historical connections may be, Nairn's current concern is far broader. It is well stated in a paragraph in Chapter IV of *The Reign of ETS*:

... throughout the odyssey of the testing profession's scientific verdicts, there have been at least two constants. First, there is the consistent finding that the working class, the poor, and the most oppressed minorities of a particular historical era have relatively little "intelligence," "aptitude," or whatever the profession is purporting to measure at the time. And, second, one consistently finds an institutional arrangement whereby the professionals who construct these tests and report these findings are backed and employed by the nation's richest and most powerful institutions; they persistently report and interpret their findings with confidence in their authority to direct the course of people's lives and thoughts (p. 162-3).

Given Nairn's broad egalitarian concern, a natural question is why he has focused his attack on ETS and its tests. An answer is suggested in a lengthy

footnote which concludes:

The practice of ranking by ETS scores and perhaps more important the institutional claim (and public belief) that this is in some sense a ranking by merit, is something which people can begin to understand and change, before they succeed in constructing a society with a new definition of economic justice (p. 454).

#### On Jensen's "Bias . . ."

If Nairn in attacking ETS as a target of opportunity engages in guilt by association, Jensen's *Bias in Mental Testing* indulges in innocence by disassociation. A minor sign of this tendency comes in the book's dedication and preface. *Bias in Mental Testing* is dedicated "to the memory of the great pioneers" of mental testing: Galton, Binet, and Spearman. In his preface, Jensen briefly reviews the contributions of these three men to the history of mental testing—but gives no attention at all to the social policies they espoused on the basis of their pioneering work in mental testing. Jensen's failure to mention Galton's espousal of eugenics seems particularly noteworthy since Jensen's own policy recommendations in the last chapter of his book are so sharply different from policies which early mental testers recommended. Jensen also refers to Sir Cyril Burt as a "great name" in mental testing (p. xii). But despite thirteen references to Burt's work, Jensen neglects to mention that much doubt has recently been cast on the quality of some of Burt's research on mental testing and strong evidence has been advanced to suggest that some of Burt's data were fabricated.<sup>6</sup>

Jensen does give attention to some criticisms of mental testing, but curiously, he does not provide references for most of the criticisms he cites. This may of course represent noblesse oblige on Jensen's part; it would be unseemly to actually identify who said such silly things.

Jensen's second chapter reviews a half dozen prominent court cases regarding testing. Though he doesn't mention the Florida case of *Debra P. v. Turlington*, he does briefly comment on the *Larry P.* case. He charges that retesting of some of the plaintiffs in the *Larry P.* (yielding IQ scores 15 to 30 points higher than previously obtained by a school psychologist) had been incorrectly conducted and complains that the "defendants did not challenge these results or raise the crucial question of whether the validity of the higher IQs obtained under . . . non-

<sup>6</sup> The fraudulence of some of Burt's work has been exposed mainly due to the skepticism of two Americans, Professor Leon Kamin of Princeton University and Professor D. D. Dorfman of the University of Iowa. The most thorough work on the quality of Burt's research is the recent biography, *Cyril Burt, Psychologist* (Cornell, N.Y.: Cornell University Press, 1979).

standard testing conditions is enhanced or reduced" (p. 33). He also observes that the defendants in the Larry P. case, namely state of California education officials, "acquiesced to the charge of test bias . . ." (p. 33).

Test critics and ill-informed defenders of standardized testing, in Jensen's view, obviously know very little about psychometrics or the science of mental testing. Getting down to serious business in his third chapter, Jensen explains:

Some of the terms in the testing controversy have become emotionally loaded and now cannot be used without misunderstanding unless they are divested of their affective overtones. Also, the meaning of certain terms in the context of psychometrics does not coincide with their meaning in common parlance or with their dictionary definition. To get on with the job, we have to agree for the time being at least on the meaning of such terms as "discrimination," "bias," and "unfair" (p. 41).

Jensen distinguishes consistently between concepts of test bias and those of unfairness:

Bias is . . . clearly distinguished from the concept of "unfairness." The concept of unfairness versus the fair use of tests is a judgment based on a philosophic position regarding the way test scores should be used, particularly in educational and employment selection (p. 454).

In other words, an unbiased test may be used in an unfair manner, and a biased test may be used in a fair way. According to Jensen, "Psychometric bias per se can be defined and identified in terms of two main classes of statistical criteria: (1) external or predictive validity criteria and (2) internal or construct validity criteria" (p. 454). There is not space here to review the various ways by which Jensen suggests how these two aspects of bias can be examined, or to recap the dozen or so definitions of fair selection which he reviews. Suffice it to say that his review of the literature on these issues is remarkably thorough, even if the perspective he brings to it is not altogether "divested of affective overtones." Hence, for the moment, let us simply skip to Jensen's overall conclusion on the bias issue, namely that

. . . many lines of psychometric evidence converge to the conclusion that, by and large, current standardized tests of general mental ability and scholastic achievement, as well as many vocational aptitude tests, are not biased with respect to any native-born English speaking minority group in the United States.

. . . In short, most standard tests, with rare exceptions, are psychometrically equivalent for minority and majority groups; the tests measure largely the same ability factors in minority and majority groups, with essentially the same reliability and validity, as these terms are traditionally defined in psychometrics (p. 715).

Jensen's conclusion is, in short, just the opposite of Nairn's. Again, we are confronted with diametrically opposed conclusions, derived from lengthy investigations of standardized testing.



### Disagreement

What is one to make of all this? The competency testing movement seeks to extend the public reach of standardized testing at the same time as the truth in testing movement seeks to curtail and control the role of tests. One federal judge in Florida rules that a standardized test in that state has not been shown to be biased or invalid while another federal judge in California rules that standardized tests of intelligence are invalid and biased. In both cases testing experts disagree sharply on issues of test bias and validity. In a lengthy treatise, Nairn argues that standardized aptitude tests are biased while Jensen in an even longer treatise argues that such tests are not biased.

This article has given only very brief descriptions of these movements, court decisions, and treatises. I have tried to provide at least a little insight into each, but far more could be said about any one. Yet perhaps more interesting than the details of any one is the question of how to make any sense at all of the very different messages they carry concerning standardized testing. Answers to this question are not, I think, to be derived from analysis of the details of any one movement, court case, or treatise. Rather,

the differences derive from deeper underlying assumptions about the relative rights and prerogatives of individuals versus those of state agencies and institutions, about the different social functions served by standardized testing, and at root from underlying differences in educational and social philosophies.

On the first point there seems to be a clear, if imperfect, correspondence between attitudes toward standardized testing and perceptions of the relative interests of individuals versus state agencies and social groups. Judge Peckham was obviously impressed with the plight of minority children placed in EMR classes, and apparently at least partly as a result put the burden of proof for the continued use of intelligence tests in such placement on the state of California; a burden which the state failed to carry out to the judge's satisfaction. In Florida, Judge Carr recognized a legitimate state interest in establishing a high school graduation requirement in the form of a test and left the burden of proof in challenging validity of the test on the individual plaintiffs—a challenge that they took up, but that failed to convince the judge. Indeed, the competency testing movement generally represents a government effort to regulate and improve schooling, while the truth in testing movement attempts to strengthen the hand of students and other individuals relative to that of institutions that screen and select among them.

In a way, Jensen appears to be inconsistent. He is a strong defender of mental tests but he concludes his book by saying:

. . . the practical applications of sound psychometrics can help to reinforce the democratic ideal of treating every person according to the person's individual characteristics, rather than according to his or her sex, race, social class, religion, or national origin (p. 740).

It is a curious closing in light of past practices in mental testing; namely the frequent inference of *group* differences regarding race, social class, and national origin on the basis of *individual* mental test scores. It would seem to be another symptom of innocence by disassociation. Throughout his book Jensen seems especially concerned with black-white differences in mental test scores and just before his final closing, disclaims:

The observed mean differences in test scores between various groups are generally not an artifact of the tests

themselves, but are attributable to factors that are causally independent of the tests. The constructors, publishers, and users of tests need to be concerned only about the psychometric soundness of these instruments and must apply appropriate objective methods for detecting any possible biases in test scores for the groups in which they are used. Beyond that, the constructors, publishers, and users of tests are under no obligation to explain the causes of the statistical differences in test scores between various subpopulations. They can remain agnostic on that issue (p. 740).

Jensen's new-found agnosticism may fit well with the genre of detached scientific scholarship adopted in *Bias in Mental Testing*, but it fails to address the social functions of standardized testing. And it is, of course, the *social* role of tests that has given rise to the current troubles over testing.

### Testing's Social Role

Students of standardized testing often distinguish among several social functions they may serve; for example, selection, placement, assessment and guidance, instructional improvement, program evaluation, and accountability functions. The current troubles over testing deal with several of these different roles of tests. The truth in testing legislation is aimed almost exclusively at postsecondary selection and admissions tests. The California court case concerned the role of IQ tests in selecting out retarded students from the general population of children. Nairn's target, ETS, derives the bulk of its revenues from selection and admissions testing programs (p. 326, 488). Minimum competency testing aims at both placement and improving instruction. More broadly, all of these deal with accountability; with various forms of social accounting by institutions.

Jensen acknowledges a range of possible uses of tests. He suggests that apart from their "necessary function in theory-oriented scientific research on the nature of human differences in abilities and other traits," tests and measurements can serve four practical roles: assessment, diagnosis, placement, and selection (pp. 45-49). He proffers some interesting advice regarding the first three functions. For example he says, "The only justification for placement is evidence that the alternative treatments are more beneficial to the individuals assigned to them than would be the case if everyone got the same treat-

---

*"But if current troubles do nothing more, they may lead to closer consideration of the extent to which our educational institutions . . . can foster development of a broader range of human talents."*

ment" (p. 46). And in his concluding chapter Jensen offers the following regarding minimum competency testing (MCT), "Although the results of MCT undoubtedly highlight a serious educational problem, I cannot see MCT as in any way contributing to the solution of the problem. It appears to me to be an unnecessary stigmatizing practice with absolutely no redeeming benefits to individual pupils or the society" (p. 724). Despite such advice on these practical applications of testing, the bulk of Jensen's concern in *Bias in Mental Testing* clearly rests on issues of selection—for higher education and for occupations.

### Conflicting Social Functions

The current troubles over tests clearly derive differently from the different social functions they serve. In terms of selecting individuals for social opportunities, dispute centers on the extent to which doing well on tests reveals a person's merit, or anything else of broader social value. In this respect defenders of standardized tests tend to emphasize the point that tests are a better and more fair means of allocating social opportunities than alternative means relied upon more heavily in the past, like money, social class, and political connections. Skeptics point out, however, that doing well on tests tends to correlate with things like social class and money, and that test scores tell little about doing well in school that could not be learned from records of past school performance. Moreover, they argue that tests tell us virtually nothing new about how well people will do in life after schooling, and often constitute a barrier to fairer allocation of opportunities to blacks and others who have been socially shortchanged in the past.

#### Writing for Educational Leadership

We welcome manuscripts about any aspect of curriculum, instruction, supervision, or leadership in education. Papers should be written in direct, readable style and be as brief as possible (five to ten pages typed double-spaced). We reserve the right to edit for brevity, clarity, and consistency of style.

References may be cited in footnotes but we suggest they be in bibliographic style at the end of the article. If that form is used, references in the body of the text should be APA style, as in: (Jones, 1979). Double space everything, including quotations and footnotes.

Please send two copies. Rejected manuscripts are not returned unless the author provides a stamped, self-addressed envelope.

What is noteworthy here is that these contrasting views of the current social role of standardized tests for selection tend to employ quite different frames of reference. Those who view the selection role of tests favorably tend to contrast the current state of affairs with the way social goods were distributed in the past—which they think was worse. Those who are skeptical of the selection function of tests tend to judge matters not in terms of past selection practices, but rather in terms of future possibilities—practices they hope will be even fairer.

For other social functions of standardized tests, these frames of reference sometimes appear to be reversed. In the Larry P. case, for example, Judge Peckham's condemnation of IQ tests appeared to be closely connected to past practices in California's EMR classes, which the judge characterized as "educationally dead-end, isolated and stigmatizing." And in the conclusion to his decision, the judge directly challenged past educational practices in California:

Whatever the future . . . it is essential that California's educators confront the problem of the widespread failure to provide an adequate education to underprivileged minorities such as the black children who brought this lawsuit. Educators have too often been able to rationalize inaction by blaming educational failure on an assumed intellectual inferiority of disproportionate numbers of black children. That assumption without validation is unacceptable, and it is made all the more invidious when "legitimated" by ostensibly neutral, scientific IQ scores. We have refused to allow the continuation of EMR policies consistent only with that assumption, and it is hoped this will clear the way for more constructive educational reform (pp. 109-110).

In the Florida case of *Debra P. v. Turlington*, past educational practices also clearly influenced the court's decision. Specifically, Judge Carr ruled in favor of the plaintiffs largely as a result of the past history of de jure racial segregation in Florida's schools. The court viewed the Florida test as a legitimate means of identifying students who needed remedial instruction even though little evidence was presented concerning present and future instructional responses to the test results. Recall, however, that Popham testified that the Florida test was not biased against black and low SES students because of his belief that those who failed it would receive future "remediation."

So perceptions of social practices connected with tests—both views of past practices and hopes for future ones—shape opinions regarding tests. This point is especially clear in the contrasting conclusions in the Larry P. and Debra P. cases regarding test validity. From a purely technical point of view there is far more evidence that the IQ tests, which Judge Peckham condemned as biased and invalid, do in fact measure something of social consequence than there

is not as much evidence with respect to the Florida test, which Judge Carr ruled was not shown to be invalid or biased.

These connections between perceptions of social practices and those of standardized tests clearly suggest that Jensen's careful distinction between test bias and unfair use of tests is artificial with respect to social functions or "practical applications" of tests. Validity and bias, after all, reside not in tests themselves but in their application. The test *Standards* point out that "validity refers to the appropriateness of inferences [drawn] from test scores or other forms of assessment" (p. 25). A wide variety of inferences may be drawn from any test score, but one acid test of what inferences are in fact drawn is how the scores are used by social institutions. In short, despite Jensen's disclaimer, it is impossible to disentangle issues of test validity and bias from philosophic positions regarding the way test scores should be used.

### Differing Aims of Education

Leaving aside broader issues of social philosophy, let me mention only the way in which different aims of education seem to be intertwined with the current troubles over testing. Three commonly acknowledged aims of education are: (1) To develop each student's educational potential to the fullest; (2) To socialize students to become productive members of society; and (3) To help channel individuals into different social roles. These aims are associated with differing schools of educational philosophy and different theories of education. They coexist within our educational system but tend to receive differing emphasis at different levels. The aim of developing all children to their full potential seems most widely recognized at the elementary level. At the secondary level, emphasis tends to be placed on citizenship, although the development and sorting goals of education are also reflected in the provision of universal secondary schooling and in the tracking practiced in most high schools. At the postsecondary level, however, the aim of sorting students into different social roles clearly comes to the fore; indeed, it is reflected in the hierarchy apparent among American institutions of higher education.

The varying emphasis on aims at different levels of education helps explain why standardized testing and the practical functions it does or does not serve are viewed so differently in different instances. In the Larry P. case, the role of IQ tests in selecting children out of regular schools into dead-end EMR classes seemed especially offensive because there is a broad social commitment to developing the educational potential of all young children. Even as staunch a defender of testing as Jensen argues that the only

*Walt Haney is Staff Director for the National Consortium on Testing Project at The Huron Institute, Cambridge, Massachusetts.*

justification for placement of students into such special classes is evidence that "alternative treatments are more beneficial to the individuals assigned to them than if everyone got the same treatment" (p. 46).

This sort of reasoning is rarely applied at the postsecondary level, however, precisely because there seems to be such universal acceptance of the social role of higher education in sorting people out into different social and occupational niches. It is this role and the part which tests play in it that seem to account for the concern of those in the truth in testing movement with postsecondary selection and admissions tests. What is worth noting in this connection, however, is that the challenge of reformers like Nairn is not just to the social role of these tests, but also to the role of institutions of higher education in sorting people into prestigious and profitable occupations. For if performance on tests tells us little about performance in life after schooling, neither does performance in institutions of higher education—college grades and academic credentials—tell us much about subsequent performance in occupations and professions. Ralph Nader's call for wider consideration of traits like perseverance, wisdom, idealism, and creativity is not just a comment on standardized tests, but more fundamentally a challenge to a social system that more and more seems to rely on test scores and educational credentials—measuring a restricted range of academic talent—to sort people into wide-ranging positions of social and economic advantage.

In the middle range of our educational system—say, the last two years of high school—there seems to be more ambivalence about the conflicting aims of sorting and development. This ambivalence is often obscured by the functions of training and socialization for adulthood, but the tension is quite apparent in the minimum competency testing movement. On one hand, minimum competency testing would be used to sort students, identifying those who are "minimally competent" from those who are not. On the other hand, minimum competency testing has been advocated as a means of developing in all students at least minimum levels of basic skills.

The discrepancies deriving from these differing aims cannot be resolved by testing experts alone—their sharp disagreement in the Debra P. and Larry P. cases nicely suggests. But if the current troubles over testing do nothing more, perhaps they will lead to closer consideration of the extent to which our educational institutions, instead of simply sorting students into social and occupational niches in the broader society on the basis of a narrow range of academic performance, can foster development of a broader range of human talents. Edward Wynne recently observed that in answering Ralph Nader's charge that tests do not measure character, a senior vice president of ETS said, "I don't see what that necessarily has to do with higher education." As Wynne wisely asks, "But if higher education doesn't have to do with character, why are we giving it our money? Just to produce good test-takers?"<sup>7</sup> *EJ*

<sup>7</sup>E. Wynne, "The College Testing Controversy," *Wall Street Journal*, 14 February 1980.



**Harvard  
University**

**Harvard Summer School  
Institute In Moral Development  
and Moral Education**

with Prof. Lawrence Kohlberg, Drs. Daniel Candee, Judy Codding, Lindsey Chesbrough, David Sachs, and Elsa Wasserman.

**July 7-August 1, 1980, M-F 9:00am-4:30pm.**

Designed to introduce teachers, counselors, and administrators to practices in the field of moral development and moral education in elementary and secondary schools. Emphasis on putting theory into practice. 8 graduate credits.

For registration information contact: Harvard Summer School, 20 Garden St., Cambridge, MA 02138. 617-495-2921.

**Center For Moral Education  
Moral Judgement Scoring Workshop  
June 23-July 2, 1980**

Workshop will be a small, intensive experience in learning to administer and score Kohlberg's moral judgement interview.

For registration information contact: Mimi McCarthy, Center for Moral Education, Harvard University, Cambridge, MA 02138. 617-495-3546  
Fee: \$200.

**Further Information on Current Testing Issues**

**Minimum Competency Testing**

Jaeger, R. M., and Tittle, C. K., eds. *Minimum Competency Testing: Motives, Models, Measures and Consequences*. Berkeley, Calif.: McCutchan Publishing, 1980. A volume of papers prepared for a special conference on minimum competency testing sponsored by the American Educational Research Association.

Airasian, P.; Madaus, G.; and Pedulla, J., eds. *Minimum Competency Testing*. Englewood Cliffs, N.J.: Educational Technology Publications, 1979. Another book of wide-ranging readings on minimum competency testing. Particularly helpful is the chapter entitled "A Citizen's Introduction to Minimum Competency Programs for Students."

**Truth in Testing**

Brown, R., and McClung, M. *Searching for the Truth About Truth-in-Testing Legislation*. Denver, Colo.: Education Commission of the States, January 1980. Contains useful review of recent truth in testing legislation, and analysis of legal issues involved. Review of nonlegal arguments is very general and sometimes superficial.

Strenio, A. *The Debate Over Open Versus Secure Testing: A Critical Review*. Cambridge, Mass.: National Consortium on Testing Project, Huron Institute. Reviews arguments both for and against open testing and evidence bearing on those arguments. Concludes with a tentative endorsement of open testing, October 1979.

Fremer, J., and Irby, A. "Why Should Some Tests be Secure?" Paper presented at National Consortium on Testing Conference, April 1979, Educational Testing Service, Princeton, New Jersey. Reviews arguments for security with respect to some tests which contribute to selection decisions.

**Recent Court Cases**

Debra P. v. Turlington. No. 78-892 c.v. T-C United States District Court N.D. Florida. July 12, 1979. 474 *Federal Supplement* 244-267. Decision in which court enjoined Florida from denying high school diplomas on basis of literacy test for four years because the test perpetuated effects of past racial discrimination and because it had not been phased in adequately. On issues of test validity and bias, court sided with the state, ruling that test had not been shown to be invalid or biased. Decision now under appeal.

Larry P. v. Riles. No. c-71-227ORFP. United States District Court N.D. California. October 16, 1979. *Federal Supplement*. Decision in which court ruled that California has used IQ tests that are racially and culturally biased for placement of students into EMR classes. On this and other grounds, court forbade the state from using IQ tests for such placement. Decision now under appeal.

**Other Topics**

Nairn, A., and associates. *The Reign of ETS: The Corporation That Makes Up Minds*. Washington, D.C.: Learning Research Project. Nader report on the Educational Testing Service attacks not only the organization, but also the validity and utility of the tests it develops.

Jensen, A. R. *Bias in Mental Testing*. New York: Free Press, 1980. Book presents a massive defense of mental tests. Argues that differences in test scores between groups represent real differences and do not result simply from bias in tests.

Fallows, J. "How Fair Are the College Boards?" *Atlantic* (February 1980): 37-48. One of the most thoughtful and best written of recent popular press accounts of recent troubles over testing.

Copyright © 1980 by the Association for Supervision and Curriculum Development. All rights reserved.