Summer 1980

Jensen and the Science of Psychometrics: A Legal Perspective - Bias in Mental Testing, by Arthur R. Jensen

Bryant Garth
Indiana University School of Law

Follow this and additional works at: http://www.repository.law.indiana.edu/ilj
Part of the Law and Psychology Commons, and the Social and Behavioral Sciences Commons

Recommended Citation
Available at: http://www.repository.law.indiana.edu/ilj/vol55/iss4/3

This Book Review is brought to you for free and open access by the Law School Journals at Digital Repository @ Maurer Law. It has been accepted for inclusion in Indiana Law Journal by an authorized administrator of Digital Repository @ Maurer Law. For more information, please contact wattn@indiana.edu.
Jensen and the Science of Psychometrics: A Legal Perspective


Bryant Garth*

An effort to confront the science of psychometrics from a legal point of view must begin with the recognition that lawyers and courts have had considerable difficulty with social science data. The general problem has stimulated a growing body of commentary, particularly in the field of law and education.¹ The marriage of law and social science envisioned by Brown v. Board of Education² has not been a happy one.³ A number of commentators, for example, have suggested that the courts have erred because they

---

* B.A. 1972, Yale University; J.D. 1975, Stanford University; Ph.D. 1979, European University Institute (Florence). Assistant Professor of Law, Indiana University School of Law (Bloomington). The author would like to thank Professor Hendrik Hartog and Dr. Susan Stager for their helpful criticisms of an earlier draft of this review.


³ Kenneth B. Clark, whose work was cited in the famous footnote 11 in Brown, id. at 484 n.11, stated:

In the optimism of the early 1950's—an optimism I not only shared but participated in with enthusiasm—social scientists, with a few exceptions, tended to accept uncritically the belief that they not only had a role but the obligation to organize, interpret, and make available to policy-makers . . . social science data that bore directly or indirectly on social issues.

have not been well-apprised of social science learning.⁴ Others have contended that judicial failures have reflected instead the inadequacy of the social science data itself.⁵ Still another group of commentators has argued that issues involving complex social science (or natural science) questions are just too difficult for regular courts ever to handle properly.⁶ The problem is a serious one, and the boom in law-related empirical research and new types of legal activism suggests that it will become even more important in future years.

The problem of reconciling law and social science is highlighted in the field of mental testing. Not only is such testing tied up with complicated issues of educational policy, but also the science of mental “measurement”—psychometrics—is an extremely technical one. Nevertheless, the law is becoming increasingly concerned with these issues.⁷ Federal legislation providing opportunities for the handicapped implicates testing issues,⁸ civil rights legislation inevitably raises the problem of testing that has an adverse impact on protected minorities,⁹ and the momentum of court-ordered desegregation has led to inquiries into school admissions¹⁰ and classifica-

---

⁵ E.g., Levin & Hawley, Foreword to Symposium, supra note 1, at 3:
Almost all of the social scientists writing for this symposium, in reviewing the existing research—published and unpublished—in their respective fields note how little competent research has actually been done to date. Accord, Yudof, supra note 3, at 108-10.
⁶ Betsy Levin, for example, emphasizes that “the adversary system process is not always the most effective way of presenting an issue.” Levin, School Desegregation Remedies and the Role of Social Science Research, Law & Contemp. Probs., Autumn, 1978, 1, 35.
⁷ For an argument that these kinds of issues should be resolved by an institution other than a regular court, see Thibaut & Walker, A Theory of Procedure, 66 Calif. L. Rev. 541 (1978).
⁸ See generally Bersoff, Regarding Psychologists Testily: Legal Regulation of Psychological Assessment in the Public Schools, 39 Mo. L. Rev. 27 (1979).
⁹ The regulations promulgated under the Education of the Handicapped Act of 1975, § 5(a), 20 U.S.C. § 1417(b) (1976), require, for example, that “[t]ests and other evaluation materials . . . have been validated for the specific purpose for which they are used.” 45 C.F.R. § 121a 532(a) (1979). A similar regulation in the rehabilitation area is found at 45 C.F.R. § 84.35(b)(1) (1978).
¹¹ The role of standardized testing in admissions to higher education has been raised by cases such as Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265 (1978), and DeFunis v. Odegard, 416 U.S. 312, 337-41 (1974) (Douglas, J., dissenting). In Bakke, Justice Powell’s opinion for the Court stated:
tions and assignments of minorities within schools. The recent Nader Report and the Nader-inspired “truth-in-testing” movement have put psychometrics further on the legal defensive. It is particularly appropriate, therefore, for legal analysts to examine the social science of psychometrics. Such an examination will be my basic purpose in this review.

Arthur Jensen’s recent book, Bias in Mental Testing, can be seen as the psychometric response to legal and other policy-oriented attacks on standardized testing. It criticizes in detail particular legal decisions, the assumptions of testing critics and the details of the arguments against testing. As Jensen claims at the outset, those who attack the tests as biased “will henceforth have this book to contend with”; it is an “exhaustive review of the empirical research bearing on this issue.” The book is meant not just for technical specialists, but also, as emphasized on its jacket, for lawyers, educators, parents and others interested in the role of

Nothing in this record—as opposed to some of the general literature cited by Mr. Justice Brennan, Mr. Justice White, Mr. Justice Marshall, and Mr. Justice Blackmun—even remotely suggests that the disparate impact of the general admissions program at Davis Medical School, resulting primarily from the sort of disparate test scores and grades set forth in n.7, supra, is without educational justification.

438 U.S. at 308 n.44.

A number of commentators have suggested that the case for special admissions programs should be based on a specific attack on the standardized tests used for admission. See, e.g., Bell, Introduction: Awakening After Bakke, 14 HARV. C.R.-C.L. L. REV. 1, 3-4 (1979); White, Culturally Biased Testing and Predictive Validity: Putting Them on the Record, 14 HARV. C.R.-C.L. L. REV. 89 (1979). See generally A. Sindler, Bakke, Defunis, and Minority Admissions (1978).


A. JENSEN, BIAS IN MENTAL TESTING (1980).

Id. at ix.
standardized testing in our society.

Jensen amasses a tremendous amount of technical data as part of his counterattack on critics such as those who have sought to put tests on trial in the courts. It is tempting for the uninitiated lawyer to despair when confronted with all this technical data. A careful examination of Jensen's scientific argument from a legal policy-oriented point of view, however, provides surprising results; it vindicates much of the recent legal involvement in testing issues. Without challenging Jensen's technical analysis, I think it can be shown that, while his defense of testing may be successful on one level, it fails to be persuasive on the major questions of interest to the law. The tests themselves may not be biased in a psychometric sense against minorities, but many normal uses of the tests lead to unjustifiably "biased" results in the sense of "unfair" denials of opportunity to members of certain minority groups. When Jensen's assumptions are distinguished from his science, and when elementary legal assumptions are factored into the analysis, his "defense" of testing turns into a fairly straightforward, limited argument implying that standardized testing should be the basis for classification and placement in a relatively small number of settings.

**The Starting Point: Equality of Potential**

The starting point for a modern legal analysis must be the assumption of equality of intellectual potential among particular racial groups and social classes.¹⁶ Consistent with our constitutional

---

¹⁶ I recognize that an argument can be made for legal adoption of Jensen's "agnosticism," at least in some cases. The assumption of equality of potential may, for example, appear to provide little guidance in cases where intentional discrimination must be shown. The key issues involved with testing, admissions, and placements, however, tend to force a choice between the assumptions of inequality and equality of potential. See notes 22-26 & accompanying text infra. In addition, the assumption of equality is implicit in the law in several areas, including the testing cases discussed in the text, school desegregation, and employment discrimination. The recent Supreme Court decision in Milliken v. Bradley, 433 U.S. 267 (1977), for example, held that remedial instruction and other compensatory education programs could be ordered as part of a desegregation remedy on the theory that it was necessary to restore victims of discrimination to the position they would have occupied in the absence of discrimination. Id. at 281-88. The implicit assumption is that the compensatory programs are necessary to raise the performance of the victims of discrimination to the level of the white majority, which is what would presumably have been the case in the absence of discrimination. Some commentators have urged that the courts infer discrimination against minorities in a racially imbalanced school from, inter alia, "scores on standardized tests," again assuming performance of groups would have been the same absent discrimination in education. See, e.g., Fiss, Racial Imbalance in the Public Schools: The Constitutional Concepts, 78 Harv. L. Rev. 564, 604 (1965). For a critical discussion of how courts
interpretation of equal protection, it would be impermissible to limit individual opportunities for minorities on the assumption that members of the group are, on the average, intellectually less capable than members of the white majority. The assumption must be the opposite one, and a heavy burden should be imposed on one who seeks to challenge that assumption.

Jensen is ambivalent about the assumption of equality. On the one hand, he signals a retreat from his statement in 1969 that “[c]ompensatory education has been tried and it apparently has failed,” primarily because of “genetic factors.”

Jensen now states:

We no longer speak of any kind of test score as a measure of a person’s capacity or potential. . . . The notions of capacity and potential suggest some clear-cut and inexorable upper limit of development. But this is a metaphysical rather than a strictly scientific notion.

Moreover,

The nearest we can come to giving any scientific meaning to the disfavored notions of “capacity” and “potential” is the concept of genotypic value. But . . . this concept has quite different meanings and implications than “capacity” and “potential,” and, furthermore, it can only be estimated from test scores probabilistically, with a rather wide margin of error for any individual.

Psychometrics will not tell us about a person’s capacity or potential. We cannot, therefore, use tests to characterize the potential or capacity of a race or social group.

On the other hand, Jensen emphasizes with similar force that:

and commentators have in the 1960’s and early 1970’s tended to turn an assumption of equality of potential into an assumption that nondiscriminatory schooling would lead to equality of result, see Yudof, Equal Educational Opportunity and the Courts, 51 Tex. L. Rev. 411 (1973). After the focus shifted to the issue of proving intentional discrimination, the assumption of equality of potential may also be relevant to a finding on intent. In Larry P. v. Riles, No. C-71-2270 R.F.P., slip op. at 86-88 (N.D. Cal. Oct. 16, 1979), appeal docketed, No. 80-4029 (9th Cir. Jan. 17, 1980), for example, the district court’s finding on intent was built in part on the state’s evident failure to make any effort to investigate a situation inconsistent with the equality assumption. See also Bersoff, supra note 7, at 93-94.

In employment discrimination litigation under Title VII of the Civil Rights Act of 1964, the Griggs formula for imposing a duty to validate assumes that minority individuals with qualifications similar to majority individuals hired for certain positions are generally capable of doing the same work, even if they score lower, on the average, on intelligence tests. See generally B. SCHLEI & P. GROSSMAN, supra note 9, at 65-131; 29 C.F.R. § 1607 (1979).

17 Jensen, How Much Can We Boost IQ and Scholastic Achievement?, 39 HARV. EDUC. Rev. 1, 2 (1969).

18 A. JENSEN, supra note 14, at 242.

19 Id. at 245.
"The assumption of equal or equivalent intelligence across all human populations . . . whatever its ideological basis, is scientifically unwarranted." He evidently supports some sort of agnostic position, which he uses to mount his critique of "tests on trial." He challenges in some detail the judicial acquiescence in that "unwarranted" assumption. What he misses, however, is that in these cases courts must have a working assumption, a starting point from which to consider the evidence, and courts necessarily will place a heavy burden on anyone who attempts to disprove the assumption of equality of potential.

A brief review of Jensen's analyses of a couple of court cases helps to reveal the significance of the assumption of equality and the problems with Jensen's call for agnosticism. Judge Wright in his famous Hobson v. Hansen decision refused among other things to allow the Washington, D.C. "tracking" system to continue. Children were assigned to the various curricular programs, or tracks, largely on the basis of standardized group intelligence tests used to evaluate the potential of each student for educational success. Black children, on the average, scored lower than did white children. The tracking system accordingly kept a disproportionate number of black children in lower tracks characterized by very limited instruction, thus guaranteeing their educational failure. Jensen of course defends the tests used to assign to the various tracks, and he argues that it does not matter if tests measure innate ability or potential. But obviously it does matter what the tests do show. The issue before the court was simple: either relatively more black children have that limited potential, or they do not. If they do not, some black individuals are being caused irreparable harm from misclassification. Tests cannot be utilized validly for these assignments unless, in Judge Wright's analysis, they measure potential or unless that potential is in fact distributed differently in different races. Given the grave consequences of misclassification and the invidious nature of any assumption of inequality, the burden had to be on defendants, and they could not prove either of these facts. Jensen's defense of the tests simply fails to address the central issue of the case.

Jensen also discusses in some detail the California cases of Di-
ana v. Board of Education\textsuperscript{23} and Larry P. v. Riles,\textsuperscript{24} which successfully challenged the use of IQ testing to place Hispanic and black children, respectively, in “deadend” classes for the “educable mentally retarded.” Here Jensen emphasizes in particular the judicial reliance on “the assumption that scholastic aptitude is equally distributed in all races, an assumption that went unchallenged.”\textsuperscript{25} Indeed, it is instructive that defense lawyers in Larry P. met twice with Jensen but still neglected to contest in court the assumption of equality.\textsuperscript{26} The assumption again had to be that educable mental retardation—which permits roughly about a fourth grade education—was evenly distributed among different racial groups. IQ test scores could not disprove that assumption, and in the absence of other persuasive evidence, the court had to invalidate testing used to perpetuate a grossly unequal education for many minority individuals.

Jensen’s personal assumptions about race and intelligence may point one way, but the law must assume otherwise. As a scientist, in fact, Jensen recognizes that the data supporting his assumption are insufficient to persuade those who begin with a different assumption. That is part of what makes his new book interesting. As will be seen, most of the serious policy arguments in the book are built on what Jensen evidently feels is the strongest scientific defense of psychometrics, the predictive validity of standardized test scores.

**The “Scientific” Rather than Genetic Defense of Mental Testing**

Jensen’s “scientific” argument in this book generally stays within the framework of the prevailing, legally-dictated assumption of equality. Notions of capacity or potential are disclaimed at


\textsuperscript{25} A. Jensen, supra note 14, at 33; see id. at 31.

\textsuperscript{26} There was evidently some disagreement between the State Superintendent of Public Instruction, Wilson Riles, who emphatically rejected any genetic hypothesis, and defense counsel. In any event, the defense in Larry P. argued primarily that racial imbalance in the classes related to nutritional deficiencies among black children, which led to an increased incidence of mild mental retardation. Noting inter alia that the incidence of severe mental retardation was not different between the black and white populations, the court rejected that contention. See Larry P. v. Riles, No. C-71-2270 R.F.P., slip op. at 43-45 (N.D. Cal. Oct. 16, 1979), appeal docketed, No. 80-4029 (9th Cir. Jan. 17, 1980).
the outset, as noted before, and they are also separated carefully from questions of testing *validity*:

Scientifically, all we can do in any test situation is measure an individual's performance then and there on the particular test at that particular time with that particular examiner. . . . From this performance measure—the test score—we can make *conditional probability* statements based on research with the particular test.27

Furthermore: “Conditional probability statements based on empirical research simply do not include or require the notions of capacity or potential.”28

The relatively narrow implications of this scientific defense of testing can be flushed out by examining briefly, with Jensen’s guidance, methods based on conditional probabilities used to establish the “validity” of standardized tests. Jensen refers to the four types of validity as “the four c’s”—content, construct, concurrent and criterion validity. “Concurrent validity,” as Jensen defines it, merits little attention here. It refers mainly to the comparison of one test with another taken at the same time, and Jensen freely admits that “[t]here are dangers in this type of validation.”29 “Content validity” findings also are outside our major concerns. That validity refers simply to the extent that test items fairly sample the appropriate universe of knowledge and “is most relevant to achievement tests, job-knowledge tests, and work sample tests.”30

“Construct validity” raises more serious issues for legal policy. This is the “attempt scientifically to understand, in psychological terms, what the test measures.”31 A careful examination of this attempt is needed, because a complete acceptance of construct validity may constitute an admission that “intelligence tests” really do measure “intelligence,” or at least give a good prediction of how “intelligent” a person is. The validation of test results in that sense could cut against our legal assumption of equal potential. Jensen, however, admits to serious limitations on construct validity and, more importantly, makes few considered pronouncements claiming that intelligence tests can effectively measure or predict any fixed status, as opposed to providing a prediction of a future performance.

27 A. JENSEN, supra note 14, at 242.
28 Id. at 243.
29 Id. at 302.
30 Id. at 297; see id. at 741.
31 Id. at 303.
Jensen notes that construct validity "is a complex, open-ended affair," and that the "task of construct validation is never really completed." It depends on the plausibility of the theoretical work being done by psychologists. That work at best provides little support for an effort to use tests to ascertain "intelligence" as a fixed and objective quality. Jensen quotes with approval, for example, the definition of intelligence provided by Lloyd G. Humphreys, another staunch defender of testing. According to Humphreys, intelligence is "the entire repertoire of acquired skills, knowledge, learning sets, and generalization tendencies considered intellectual in nature that are available at any period of time," with the skills that are "intellectual in nature" taken from a "consensus among psychologists at any given time." Even if we accept the argument that intelligence tests accurately tap this "repertoire of acquired skills," that does not solve the problem of what low "intelligence" means in terms of the ultimate potential of a person to achieve. It can only give rise to predictions of future performance, which are discussed below. Jensen therefore does not argue that construct validity justifies firm conclusions based upon the results of intelligence tests.

In one instance, however, Jensen appears to claim that intelligence tests can measure capacity to learn. Early in the book he defines "mental retardation" parenthetically as IQ’s below 70. Subsequently he argues that a low IQ "is disabling in mathematics . . . and in English composition and reading comprehension" and claims that a child with an IQ of about 75 or below "cannot master the traditional subject matter of elementary school." If this assertion were accepted, it would appear to justify the treatment of children with low IQ’s as lacking the potential to master certain skills. It appears, however, that Jensen is merely reiterating a con-

---

22 Id. at 420.
23 Id. at 304.
24 Id. at 170 (quoting Humphreys, Theory of Intelligence, in INTELLIGENCE 31, 31-32 (R. Cancro ed. 1971)).
25 Id.
26 Jensen’s long explication of the hypothetical construct termed g represents an effort to equate intelligence with what the tests measure. See A. JENSEN, supra note 14, at 169-258. But even if g exists as postulated by Jensen, it does not necessarily require one to accept the notion of innateness or limited potential. At best, it represents a current ability to perform on a number of tests.
27 Id. at 84.
28 Id. at 107.
29 Id. at 114.
venient, overly simplified, psychological definition. Significantly, Jensen rather uncharacteristically cites no studies to support the permanently "disabling" nature of a low IQ, and also uncharacteristically provides no comparative racial and ethnic group data. It is perfectly possible, for example, that IQ scores could predict future performance accurately for black and white children, but that black children with the same scores as middle class whites might achieve better if given a quality education or a home environment more conducive to normal school learning. Jensen's asides about the meaning of low IQ's must be seen as outside his central argument in this book.

The key to Jensen's scientific defense of testing is the ability of tests to predict future performances—"criterion" or "predictive" validity. As he defines it, "[t]his is the ability of test scores to predict performance in some endeavor that is external to the test itself, called the criterion." Its central importance, and the reason it cannot be neglected, is that it "depends entirely on empirical demonstration." It "is probably the most important, defensible, and convincing type of validation in the practical use of psychological tests." Predictive validity, as will be seen, is not inconsistent with the assumption of equal potential.

The discussion of predictive validity in the book may be subject to some methodological challenge, especially as it relates to minority groups. Jensen admits, for example, that for elementary school "the published evidence . . . is surprisingly meager," and, for secondary school, "[h]ere, too, the evidence is surprisingly scant." But Jensen does produce a substantial body of evidence for his position that "differential validity for the two racial groups [black and white] is a virtually nonexistent phenomenon." We must consider the implications of the not surprising finding that the tests do have essentially the same predictive validity regardless of

---

40 The relatively arbitrary nature of this definition is also evidenced by the history of the American Association on Mental Deficiency's definition of mental retardation. The Association had previously considered children who scored one standard deviation below the mean on intelligence tests as mentally retarded. However, in 1973 the Association revised its definition to include only those who scored two or more standard deviations below the mean. AMERICAN ASSOCIATION ON MENTAL DEFICIENCY, MANUAL ON TERMINOLOGY AND CLASSIFICATION IN MENTAL RETARDATION 4, 11 (1973 rev.).

41 A. JENSEN, supra note 14, at 298.
42 Id.
43 Id.
44 Id. at 472.
45 Id. at 474.
46 Id. at 515.
Moreover, it is notable that recent criticisms of testing, such as the Nader Report on the Educational Testing Service, do not focus on any differences in predictive validity between different races or ethnic groups. Jensen is correct in his argument that predictive validity is the most convincing standard by which to defend the tests, and we must consider the implications of that type of validity.

**Implications of "Predictive Validity"**

Predictive validity of intelligence tests is the central claim of the psychometricians. As Jensen describes it, the validity varies somewhat among various standardized intelligence tests. Tests designed specifically to measure intelligence, such as the Wechsler battery of tests and the Stanford-Binet, tend to correlate with achievement by .60-.70 in elementary school, .50-.60 in high school, .40-.50 in college, and .30-.40 in graduate school. As the declining correlations indicate, other standardized tests tend to have a greater validity as individuals get older. Such tests can measure specific learning—achievement—which "involves more different identifiable causal factors and correlates than IQ, which is simply the single most important factor." Achievement relates to other personality traits, such as motivation, emotional stability, persistence and work habits.

Tests such as the Scholastic Aptitude Tests, which combine achievement and ability, predict grades in college by some .30-.70, with a mean of .50. A test such as the Law School Admission Test—"essentially a high-level verbal intelligence test"—correlates about .30 with first year law school grades. The aptitude and intelligence distinction, however, is not a very clear one. Jensen quotes with approval a statement that "the quality of students admitted to a graduate program would be about the same if the test used for selecting them were the GRE, the Miller Analogies Test, the

---

47 As one reviewer of Jensen has noted, "It would be a poor reflection indeed on the technical competence of psychometrics if, after nearly a century of effort, they had found no way to eliminate such an elementary and undesirable effect as differential predictive validity." Gould, *Jensen's Last Stand*, N.Y. Review Books, May 1, 1980, at 38, 39.
49 A. JENSEN, supra note 14, at 319.
50 Id. at 317.
51 Id. at 329.
52 Id. at 332-33.
the CEEB-SAT, the ACT, the Law School Admission Test, the Medical College Admission Test, or some other admission test. . . . Scores from each of these tests correlate very highly [perhaps .80] with scores from the others." All these tests are measuring precisely the same skill, however acquired. According to Jensen, that skill helps predict future educational performances among all strata of American society.

But that evidence obviously does not take us very far. As Jensen admits, the correlations with future performance are not particularly high. At one point, for example, he states that, in order to predict 50 percent better than chance, it would take a correlation of .866. Another way to see the same problem is by calculating "variances" accounted for by test scores. Variance, meaning here the proportion of the total variation in one variable (e.g., academic performance) that is predictable from another variable, is obtained by squaring the correlation coefficient. In other words, to account for the percentage of future academic performance by an individual predictable from the SAT alone, we can square .50, obtaining .25. That means roughly that we can predict an individual's performance with a margin of error that is only 25 percent smaller than it would be if we were to guess.

Furthermore, the predictive validity of intelligence and aptitude tests for older students, as noted before, tends to lag behind the simple indicator of past academic performance. While the "subjectivity" of grades is emphasized by Jensen, he admits that grades do have predictive validity among older students that is as good or better than that provided by the tests. For example, "[h]igh school grades or the student's rank in his or her graduating class generally predict college GPA at least as well as scores on college aptitude tests."

The phenomenon of the decreasing educational predictive validity of general aptitude or ability testing of older individuals is duplicated in the employment area. Jensen thinks there is a threshold IQ for certain occupations, but it is of course difficult or impossible to determine that threshold. In general, however, "[t]he IQ and other ability test scores are considerably better at predicting persons' occupational statuses than at predicting how well they

---

53 Id. at 469.
54 Id. at 305.
55 Id. at 255.
56 Id. at 330.
57 Id. at 363-64; see id. at 347-53.
will perform in the particular occupational niche they enter.” Experience, dedication, reliability, diligence, ability to get along and other personal factors again become more important than IQ in ultimate occupational success.

Serious limitations on standardized testing as described by Jensen therefore include a relatively low predictive validity for future educational or occupational performance, and the fact that other indicia, above all past performance, tend to have better correlations with future success. To these limitations must be added a more fundamental one that is implicit in any reliance on predictive validity. Predictive validity tells you that, with no change or intervention, it is more or less likely that the existence of one factor will accurately predict the existence of another. To base individual decisions on predictive validity, however, is to assume that changes cannot, or at least will not, be made to increase the likelihood of a desired result. More concretely, remedial education, individualized tutoring, on-the-job training, and other programs may make the test score’s prediction fail. To the extent that such programs exist and can succeed, and Jensen here expresses no final opinion on that point, obviously we should rely less on test scores. To the extent special programs do not exist and have not been tried, we can question whether permitting reliance on test scores is simply an impermissible excuse to perpetuate and legitimate inequality. A further problem, however, is that in many instances the costs of closing an educational or training gap may be deemed prohibitive. Even if it is possible to take an illiterate forty-year-old and make him a competent lawyer, we probably would be hesitant to expend the resources required to do the job, whereas remedial reading for eight-year-olds may be relatively easy to justify.

Jensen recognizes a number of the problems raised by reliance on predictive validity. To his credit he examines some of the policy implications which seem naturally to follow from a more qualified acceptance of predictive validity, even though their implementation would result in a cutback in mental testing. For children in primary schools, for example, Jensen argues against the administration of intelligence tests despite the fact that these tests predict educational performance best among young children. He rejects testing at young ages because the likely use of such tests is to limit irrevocably the opportunities available to those with low scores (who will often be poor or from minority groups). Intelligence

88 Id. at 347.
tests, as the court cases discussed earlier showed, have been given traditionally for the purpose of "ability grouping," typically including a "track" analogous to that sometimes called "educable mentally retarded." Children in such tracks are generally deemed (consistent with Jensen's assertion quoted earlier) to be incapable of mastering a normal grade school curriculum. If "errors" are made in assignment to this group, those who are misassigned are subjected to a grossly inferior, permanently limiting education. If we start with the assumption of equality and recognize the margin of error in intelligence tests at their best, we should be reluctant to tolerate discriminatory impacts here. Indeed, on the basis of the margin of error alone Jensen reaches the same conclusion through a kind of "burden of proof" analysis:

There is no compelling evidence that would justify ability grouping in the elementary grades. Grouping into separate classes on the basis of either IQ or achievement test scores not only stigmatizes the pupils in some classes as slow learners, but limits the educational aspirations and opportunities of those children placed in the slow groups, making it still more difficult to catch up or keep up with their agemates if they are capable of doing so.

Jensen's scientific argument thus propels him to a conclusion consistent with Hobson v. Hansen and Larry P. By eliminating his assumption of inequality, he has moved to a position consistent with the assumption of equality.

For older children Jensen's awareness of the limits of his science again causes him to be reluctant to limit opportunity on the basis of test scores. For children seeking admission to "advanced" classes, he counsels:

[w]here there is any doubt about the pupil's prerequisite ability and the probability of his or her success in an elected course, I believe the best "test" is the student's performance in the course itself. The margin of error based on predictions of future performance from test scores of any kind are [sic] large enough that, when at all feasible, I would let the motivated student try, even if he then fails, rather than tell him that he should not try because of his low score on a test.

---

60 A. JENSEN, supra note 14, at 719. When Jensen discusses "special education," he does not refer to these lower tracks but rather to classes for persons with clear physical or mental handicaps. See id. at 720.
61 Id. at 720.
Jensen, therefore, suggests that test scores should not be utilized to limit opportunity in primary or secondary schools. Even if test scores are psychometrically valid in terms of predictive validity, the validity is too limited to deny individual opportunities to so-called “overachievers” or potential “overachievers.” He strikes a logical balance against the harmful use of testing, a balance that should be even easier to strike if the problem of adverse impact on minority groups is considered.

For similar reasons, Jensen believes achievement tests should be substituted generally for intelligence tests: “There is no routine use of IQ or aptitude tests in the schools for which well-designed scholastic achievement tests would not better serve the same purpose.” The single score of intelligence tests, with the inevitable implication of a fixed measure of potential, is simply not worth the risks of abuse. Even if school psychologists give the test in order to try to understand an individual’s academic problems, the score should not be communicated to teachers, parents, or pupils:

Description of the pupil’s educational problem should be given in terms of scholastic performance per se, and its trend over the pupil’s years in school. It is part of the school psychologist’s job to try and find, through examination, any factors in the school or home situation that might feasibly be altered to the benefit of the pupil’s scholastic progress, personal development, and self-esteem.

Again, the predictive validity of intelligence tests is not such that it can or should be used to perpetuate the inequalities reflected in different scores. Primary and secondary school students can have their accomplishments measured by grades and achievement tests that cover the material supposed to be learned, but the psychometric properties of IQ tests militate against using such scores to limit opportunities for educational success or avoid developing “factors in the school or home situation” that can help beat the predictions. From a legal point of view, therefore, psychometrics should produce no hesitation for invalidation of intelligence tests that, as a matter of fact, do limit educational opportunities for members of minority groups in primary and secondary schools. Even Jensen strikes the balance against testing, doing it simply on individualistic and psychometric grounds.

But the question now is how far to go with this analysis. Jensen

\*\* Id. at 716.
\*\* Id.; see id. at 722-23.
\*\* Id. at 723.
does not strike the balance in the same way for testing as a means to screen admissions to college: "No injustice is done by the use of standardized academic aptitude tests for college admission." Yet his argument takes account of the limitations on individual opportunity inherent in decisions based on predictive validity. He states that no one is hurt by institutions that rely on the limited predictive validity obtainable from test scores. The reason is that an individual can find a college that is best suited to the predicted academic performance, and if the individual does well in that setting in the first year or two, he or she can transfer to "an academically more demanding college."

This picture of intercollege mobility is probably a bit exaggerated, but for present purposes the interesting point is that Jensen again is compelled to compensate for denials of opportunity based on test scores. Arguing on individualistic, psychometric grounds, Jensen comes to a position consistent with what might be urged from the point of view of an assumption of equality of potential.

Jensen's analysis, however, does not help us with the problem of admissions to professional schools such as those required for careers in law and medicine. There we cannot contend that everyone can find a place in some appropriate institution. A 1977 study found, for example, that only 43,513 of the 76,061 applicants to law school were offered one or more places of admission. For black candidates, the figures were 1,697 out of 4,299, and for Chicano candidates 510 out of 1,085. Many students simply do not get an opportunity to enter these professions.

Nevertheless, Jensen does not worry about denials of opportunities to these schools. Before discussing his position, two other relevant concerns that he introduces can help provide further perspective. The first concern is evinced with respect to testing generally for employment purposes. Here he emphasizes "the distinction between the prediction of performance during job training and performance after the completion of training," and he would include the initial job performance within the period of training. This distinction is essential because training and initial job performance tend to be predicted better by standardized tests than is subse-

---

66 Id. at 726.
67 Id.
67 CARNEGIE COUNCIL ON POLICY STUDIES IN HIGHER EDUCATION, SELECTIVE ADMISSIONS IN HIGHER EDUCATION 188 (1977).
66 Id. at 189.
68 A. JENSEN, supra note 14, at 728 (emphasis omitted); see id. at 312-13.
quent job performance. What this means is minorities or others who may be excluded on the basis of low test scores could become, after a period of time on the job, just as good employees as those not excluded. Therefore, in Jensen's words: "The whole issue . . . hinges on the duration and cost of the training required for a particular job, as well as on the ultimate success-failure ratio of trainees as a function of their test scores." And, for him, the idea of such training would be "wholly unrealistic . . . for those 'high-level' jobs in which not only the training for the job but the performance of the job itself constantly makes demands of a highly g-loaded [i.e., intellectual] nature."\[71\]

We do not know how much training can overcome a lack of academic skills, or at what cost, for many occupations. What is significant at this point, however, is simply to note the concerns Jensen's statement reflects. Jensen's sensitivity to the limits of predictive validity in individual cases has thus far led him to reject tracking of elementary school students on the basis of predictions of academic performance, to advocate that students generally be permitted to try and fail rather than be denied an opportunity because of a prediction of failure, and to ask whether special training efforts and experience at developing certain skills can substitute in many instances for decisions based on predictions. Predictive validity, therefore, is of limited practical meaning; it is not necessarily a license to select a person only, or primarily, on the basis of test scores that predict future success.

One further aspect of predictive validity must be examined before a more concrete evaluation of the uses and abuses of testing from a legal policy standpoint can be attempted. Jensen makes the argument that the purpose of admissions tests like the LSAT is not to predict individual performances, but to maximize the academic potential from an institutional point of view. More precisely, the utility of testing for placement in a position open only to some applicants depends on the test's "accuracy of predicting success or failure."\[72\] Given some predictive validity coupled with a selective placement process, the use of the tests can substantially decrease the failure rate of those who are accepted.\[75\] The greater the selectivity, the more useful a test with a given predictive validity.

Several implications follow from this type of defense of testing

\[70\] Id. at 728.
\[71\] Id. at 728-29.
\[72\] Id. at 306; see id. at 47-48.
\[75\] See id. at 307-09.
for placement or selection. The need to avoid limiting opportunities for individuals who are predicted to fail has already been considered. Predictive validity is too low to enable accurate prediction of the future performance of particular individuals. In addition, the success or failure rate of the group admitted, which can be predicted with some accuracy, must be seen as a somewhat arbitrary figure. This figure depends on the number of applicants and number of openings, given a certain predictive validity. The institution can to some extent adjust those factors. But an individual should be able to make no claim of right on the basis of that individual’s predicted performance. There is no objective psychometrical reason to insist on maximizing the success rate. Maximizing institutional academic success is one of many possible institutional goals, and at the very least it must be balanced against the fairness of the process to individuals. Therefore, as a matter of legal policy, for example, if the impact of a given admissions system based solely on predictive validity is adverse to minority individuals, it should not necessarily be immune from legal attack on the basis that the success rate was being maximized from an institutional perspective. The institution—Jensen’s analysis suggests—must defend its choice of the particular predicted success or failure rate.

The soundest claim of psychometrics—predictive validity—provides no compelling basis to follow one selection policy as opposed to another. Indeed the limitations of predictive validity require, even for Jensen, that it be balanced in some manner against denials of individual opportunity. If this is so, however, the question is why Jensen and others still want individuals to be ranked and selected primarily on how well they are predicted to do, given their grades and their test scores. The reason must be that they want something more than predictive validity. Their belief is that a “meritocracy” is the best possible society, and test scores help to build such a society. From a societal standpoint, a reliance on the formula for predicted academic success can maximize the success rate. But that of course is simply one form of social planning that sacrifices some qualified individuals in pursuit of a plan. It can have no more philosophical validity than a form of social planning that tries to diminish poverty and racial prejudice on the assumption that a meritocracy at this stage of our society is really a veiled

---

76 See, e.g., A. Jensen, supra note 14, at 391-98.
means of perpetuating undeserved inequality.

The above observation returns us to the assumption with which this review began. If the current distribution of wealth, power, and opportunity is basically a correct one, and minorities have failed to share proportionately in successes essentially because of their own inherent deficiencies, then a "meritocracy" that will perpetuate the status quo might be chosen. But if we assume equality of potential and look at the position of certain groups in our society, we ought not to acquiesce in the adverse impact on minorities of "objective" standards that, by Jensen's own admission, will unfairly deny opportunities to qualified individuals.

THE EFFORT TO IGNORE PREDICTIVE VALIDITY

Even if tests have a relatively limited predictive validity, however, that does not mean we should eliminate them. Yet the recent Nader Report on the Educational Testing Service suggests that the tests as presently designed should not be used at all in college and graduate admissions decisions.76 There is much useful debunking in the Nader Report,77 but its attack on standardized testing ignores the utility of instruments that can maximize our ability to predict academic performance. The Nader Report argues that particular test items are biased, that there is a majority culture reflected in the standardized tests, that the tests do not predict subsequent job performance as well as they do initial academic performance, that academic grades are by and large better predictors of academic success than are the test scores, and that the test scores are in general "no better than chance" predictors of an individual's subsequent performance.78 While the first two criticisms may have some cogency if tests are sought to be used to measure intellectual capacity or potential, Jensen argues convincingly that they are irrelevant to the question of predictive valid-

76 See A. Nairn, supra note 12, at 384-89.
77 Interestingly, Jensen makes the following criticism of the large test publishers, such as Educational Testing Service and the Psychological Corporation:

"Today we possess the necessary psychometric technology for producing considerably better tests than are now in popular use. The principal hindrances are copyright laws, vested interests of test publishers in the established tests in which they have already made enormous investments, and the market economy for tests.

A. Jensen, supra note 14, at 138.
78 See generally A. Nairn, supra note 12, at 55-160.
The other criticisms have some power, as I have suggested, although the “no better than chance” formulation is misleading in terms of what the tests can do. Still, even if grades are better predictors in many instances, that does not mean that we should not try to combine grades with test scores and other information to yield a still better predictor.

In the absence of some type of affirmative action program, there is no question that a formula for predictive validity that includes a standardized test score will have a serious adverse impact on minorities. On the other hand, the substitution of undergraduate achievement for that formula could be just as bad, since grades would probably be discounted or appreciated depending upon the relative prestige or “rigor” of the undergraduate institutions. This would be necessary to cope with the likely large number of candidates who will look similar on paper. Other substitutes such as personal or academic references tend also to carry more weight to the extent they are from conventionally successful persons, and that, too, may give an edge to the well-to-do. The problem is that underprivileged minorities tend to have worse “credentials” than middle class white students, and eliminating one credential will not affect that social fact. And if we try to construct a formula that predicts successful performance later in life, as the Nader Report suggests might be desirable, we might end up with the same formula that enabled the Princeton class of John D. Rockefeller III to select him as “Most Likely to Succeed.”

Furthermore, it is naive to ignore test scores that have some predictive validity. These scores can predict objectively that, all things being equal, if a given number of persons with these records are admitted, one should expect a certain rate of failure. This tells us that something must be done to help those who are predicted to do badly, since otherwise the prophecy will come true. To ignore predictive validity will not change academic results. Indeed, the use of some formula as a guide to admissions of “low predictors” can help the institution’s affirmative action program succeed with the minorities who are admitted specially. We should not try to hide the fact that in our present society, certain traditionally disadvantaged groups have not received the academic benefits of a middle class culture, a middle class income, and middle class

79 See A. JENSEN, supra note 14, at 635-714.
80 Compare my formulation in the text accompanying note 55 supra. This criticism is of course made in some detail in the reply of ETS to the Nader Report, A. NAIRN, supra note 12. See EDUCATIONAL TESTING SERVICE, TEST USE AND VALIDITY 16-21 (1980).
schools. That is the same as expecting academic miracles from de-
segregation. Removing a few obstacles to advancement does not by
any means overcome the fundamental problems created by poverty
and racial prejudice. Affirmative action programs can help in a
modest way. But we must admit that, all things being equal, those
who come from disadvantaged backgrounds in our society are not,
on the average, going to defy the predictions. Indeed, to the extent
tests do have predictive validity and predict poor performances for
minorities, that is a serious criticism of the system that has pro-
duced that phenomenon. The situation needs to be addressed, not
ignored.

TESTING AND THE LAW: A CONCLUSION

This review has sought to initiate and help structure a dialogue
between the science of psychometrics as presented
by Jensen and
legal policy. Beginning with what seems to be a reasonable legal
assumption and focusing on Jensen's "scientific" as opposed to
"ideological" opinions, Jensen's detailed, 800-page defense of test-
ing has been turned into a modest and useful document for those
concerned with the role of standardized testing in denying or en-
hancing educational opportunities. Seen in this manner, Jensen's
careful and limited defense of testing suggests that the law should
have little hesitation in intervening in what may otherwise have
appeared to be an area too complex for serious legal scrutiny.\(^1\)
The key to a defense of testing is predictive validity, and we must
take predictive validity seriously. Tests with predictive validity can
be important and useful in documenting what is likely to occur,
but they cannot vindicate the applications of test scores to deny
educational opportunities. Complicated issues remain about what
legal policies should be developed pursuant to the ambiguous legal

\(^1\) In the very recent case of Parents in Action on Special Educ. v. Hannon, No. 74-C-3586
(N.D. Ill. July 7, 1980), the district court judge rejected an attack on IQ tests used for
assignment to special classes for the "educable mentally handicapped." The decision up-
holding the tests was based on an analysis of each item of the tests to determine if it was
culturally biased, and the judge found one item on the Stanford-Binet and eight on the
Wechsler tests to be biased. Jensen's own defense of testing, however, suggests that the
judge's focus was on the wrong question. The tests result in a heavily disproportionate en-
rollment of black students in special classes that inevitably presuppose a limited potential
for learning, and the question must be whether the tests are valid for assignment of minori-
ties to those classes. Cultural bias against minorities does not mean that particular items are
bad but that the use of these tests limits opportunities for minorities without a satisfactory
explanation as to why that reflects an inability to take advantage of greater opportunities.
mandate for equal educational opportunity. Recent efforts in the law to scrutinize testing practices carefully should not be deterred from further efforts by the findings of Jensen’s "exhaustive review."