

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Social and Technical Issues in Testing:
Implications for Test Construction and Usage

Buros-Nebraska Series on Measurement and
Testing

1984

5. Social and Legal Influences on Test Development and Usage

Donald N. Bersoff

John Hopkins University, donald.bersoff@drexel.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/burostestingissues>



Part of the [Educational Assessment, Evaluation, and Research Commons](#)

Bersoff, Donald N., "5. Social and Legal Influences on Test Development and Usage" (1984). *Social and Technical Issues in Testing: Implications for Test Construction and Usage*. 7.
<https://digitalcommons.unl.edu/burostestingissues/7>

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Social and Technical Issues in Testing: Implications for Test Construction and Usage by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

5

Social and Legal Influences on Test Development and Usage

Donald N. Bersoff
Ennis, Friedman, Bersoff & Ewing
Washington, D.C.
University of Maryland School of Law and
The John Hopkins University

It was the Chinese over 3000 years ago, not the Americans in this century, who first used large-scale psychological testing (Dubois, 1966). But, as with many other technological developments, it was the United States that enthusiastically adopted the method (Haney, 1981). By now it is highly probable that every person in our country has been affected in some way by the administration of tests. Testing has become the means by which major decisions about people's lives are made in industry, education, hospitals, mental health clinics, and the civil service.

Tests themselves, by and large, are facially neutral. They do not inherently discriminate against those who take them and, undoubtedly, scores derived from tests have been used to admit, advance, and employ. For most people, however, test results have served as exclusionary mechanisms—to segregate, institutionalize, track, and deny access to coveted and increasingly scarce employment opportunities.

At one time, the work of academic and applied psychometricians went virtually unexamined by the law, but as the use of tests increased in the United States, so did their potential for causing legally cognizable injury to test takers. As a result, there is probably no current activity performed by psychologists so closely scrutinized and regulated by the legal system as testing.

SOCIAL INFLUENCES

Although recent litigation and legislation directly affect the continued administration of psychological testing, most especially in employment and educational settings, it is my contention that what appears to be an antitestng movement in the courts and in Congress is not an antitestng movement at all. It is my thesis that, in the main, the law's concern about testing has been evoked by the following three major social developments.

1. Our society in the last 30 years has made attempts, albeit unevenly, to undo the effects of history of *de jure* segregation and *discrimination against racial and ethnic minorities*. Many of the more familiar cases, such as *Larry P. v. Riles* (1979) affecting individual intelligence scales, *Debra P. v. Turlington* (1981)¹ concerning minimal competency tests, and *Teal v. Connecticut* (1982) litigating nuances of employment selection assessment, flow inexorably from *Brown v. Education* (1954) and are simply renewed claims by minorities for the fulfillment of the meaning of the 14th Amendment's equal protection clause. They reflect the most recent challenges to practices that are perceived as attempts to continue, in a more sophisticated manner, the racial and ethnic separation more blatantly used in the early 1950s and 1960s by educational institutions and public and private employers.

2. The courts have recognized, as a constitutional imperative, the *right against impermissible intrusion by the government* into the private lives of its citizens. Defining the right to privacy has been difficult for the courts, but recently the Supreme Court noted that one aspect of the right "is the individual interest in avoiding disclosure of personal matters [*Whalen v. Roe*, 1977, p. 598]" or as the late Justice Brandeis more esthetically phrased it, "the right to be let alone [*Olmstead v. United States*, 1927, p. 478]." If, as Reubhausen and Brim (1965) assert, the "essence of privacy is . . . the freedom of the individual to pick and choose for himself the time and circumstances under which, and most importantly, the extent to which, his attitudes, beliefs, behavior and opinions are to be shared with or withheld from others [pp. 1189–1190]," one can easily see why the broad spectrum of testing, but particularly personality and attitude testing, would be the object of legal scrutiny.

3. Finally, there has been a third social development that has influenced the law's concern with testing. Unlike judicial declarations concerning discrimination and privacy, this last influence is not of recent vintage. As a social phenomenon it has been part of human culture since its beginning. If you will pardon a highly technical psycholegal term, I would like to call this third aspect *stupidity*. Stupidity may be defined as negligence or, alternatively, the failure to use reasonable care in carrying out one's obligations. Although it does not connote

¹Complete citations for all cases referenced in the text are found in Table 5.1.

intentional or willful desire to harm, negligent harm can be just as damaging as purposefully inflicted injury. It is my thesis that stupidity, more than modern interpretations of equal protection and privacy, has been responsible for the increased legal regulation of psychological testing. On this count, both psychologists and judges must be faulted.

I believe that almost all the important legal decisions concerning psychological testing may be viewed as various combinations of the social phenomena identified as items 1 to 3 just given. I would like to spend some time in developing this thesis by giving several pertinent examples from educational, employment, and forensic settings.

SOCIAL INFLUENCES ON LEGAL DECISIONS: SOME EXAMPLES

Education

The Supreme Court's ringing declarations in *Brown v. Board of Education* (1954) ended state-imposed segregation in the public schools. But in the decade after *Brown*, many southern school systems refused to accept the Court's decision as final. They interpreted the Court's assertion that separation of black children from white "solely because of their race generates a feeling of inferiority . . . that may affect their hearts and minds in a way unlikely ever to be undone [p. 494]" as an empirically testable hypothesis, not a normative legal principle. Thus, in the early 1960s one of Georgia's school systems sought to disprove what it believed to be an erroneous factual premise. It alleged the segregation they were accused of perpetuating was not based on color "but rather upon racial traits of educational significance as to which racial identity was only a convenient index [*Stell v. Savannah-Chatham County Board of Education*, 1963, p. 668]." They attempted to show that differences in learning rates, cognitive ability, behavioral traits, and capacity for education in general were so great that not only was it impossible for black children and white children to be educated effectively in the same room but that to "congregate children of such diverse traits in schools . . . would seriously impair the educational opportunities of both white and Negro and cause them grave psychological harm [p. 668]."

To prove their contentions the defendants called several expert witnesses, among them two psychologists, Travis Osborne and Henry Garrett. Based on such instruments as the California Achievement Test and the California Mental Maturity Tests, they testified that significant differences in test scores were indicative of inherent differences in the races and that only minor changes could be achieved by educational readjustment or other environmental change. Al-

TABLE 5.1
Table of Cases

Albemarle Paper Co. v. Moody, 422 U.S. 405 (1975).
Ballew v. Georgia, 435 U.S. 223 (1978).
Battie v. Estelle, No. 79-1567 (5th Cir. Sept. 11, 1981).
Brown v. Board of Education, 347 U.S. 483 (1954).
Debra P. v. Turlington, 644 F.2d 397 (5th Cir. 1981).
Estelle v. Smith, 451 U.S. 454 (1981).
Firefighters Institute v. City of St. Louis, 616 F.2d 350 (8th Cir. 1980), <i>cert. denied sub nom.</i> United States v. City of St. Louis, 452 U.S. 938 (1981).
Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).
Griggs v. Duke Power Co., 401 U.S. 424 (1971).
Guardians Association v. New York Civil Service Commission, 630 F.2d 79 (2d Cir. 1980), <i>cert. denied</i> , 452 U.S. 939 (1981).
Hobson v. Hansen, 269 F. Supp. 401 (D. D.C. 1967) <i>aff'd sub nom.</i> Smuck v. Hobson, 408 F.2d 175 (D.C. Cir. 1969).
Kirkland v. New York State Dep't. of Correctional Services 520 F.2d 430 (2d Cir. 1975), <i>cert.</i> <i>denied</i> . 429 U.S. 823 (1976).
Larry P. v. Riles, 495 F. Supp. 926 (N.D. Cal. 1979) <i>appeal docketed</i> . No. 80-4027 (9th Cir., Jan. 17, 1980).
Merriken v. Cressman, 364 F. Supp. 913 (E.D. Pa. 1973).
Olmstead v. United States, 227 U.S. 439 (1928).
PASE v. Hannon, 506 F. Supp. 831 (N.D. Ill. 1980).
Stell v. Savannah-Chatham Board of Education, 200 F. Supp. 667 (S.D. Ga. 1963), <i>rev'd</i> 333 F.2d 55 (5th Cir. 1963), <i>cert. denied</i> , 379 U.S. 933 (1964).
Teal v. Connecticut, U.S., 102 S. Ct. 2525 (1982).
Washington v. Davis, 426 U.S. 229 (1976).
Whalen v. Roe, 429 U.S. 589 (1977).

though these test results and testimonies went unchallenged by attorneys fighting to enforce desegregation, the idea that such devices could measure innate ability found its way into a 1967 decision that, at the time, became the most persuasive and widely quoted legal opinion of its kind. That case is *Hobson v. Hansen*.

At issue in *Hobson* was not psychological testing but rather the constitutionality of disparities in the allocation of financial and educational resources in the Washington, D.C., public school system that, it was claimed, favored white children. Also at issue was the overrepresentation of black children in lower, and white children in upper, ability groups. But, in the course of the trial, it was adduced that the method by which track assignments were made depended almost entirely on such standardized group ability scales as the Metropolitan Readiness and Achievement Test and the Otis Quick-Scoring Mental Ability Test. *Hobson*, when read in its entirety, represents the justified condemnation of rigid, poorly conceived classification practices that negatively affected the educational opportunities of minority children and led to permanent stigmatization of blacks as unteachable. But swept within *Hobson's* condemnation of harmful classification practices were ability tests used as the sole or primary decision-

making devices to justify placement. Not only was ability grouping as then practiced in the District of Columbia abolished, but tests were banned unless they could be shown to measure children's innate capacity to learn. No psychologist who has written on the subject, including Jensen (1969, 1980), believes that tests solely measure hereditary endowment (Anastasi, 1976; Cleary, Humphreys, Kendrick, & Wesman, 1975). No test could pass such a criterion.

Left unscathed in *Hobson* were the stately, revered, and venerated devices against which all other tests were measured—the individual intelligence scales. But that was soon to change as the result of actions brought in San Francisco and Chicago. Two diametrically opposed decisions, *Larry P. v. Riles* (1979) (the San Francisco case) and *PASE v. Hannon* (1980) (the Chicago case) are seen by psychologists as attacks on IQ tests. That, however, is a significant misperception. Like *Hobson*, these two pieces of litigation are actually challenges to educational practices deemed to be discriminatory. [Similarly, the recent attack on minimal competency tests, see *Debra P. v. Turlington* (1981), is more appropriately seen as a claim by black children that the use of such tests is merely a subtle but effective effort by states to resegregate the public schools.]

The real issue was the basis for disproportional placement of black children in segregated, self-contained classes for the educably mentally retarded. Throughout his opinion, Judge Peckham in *Larry P. v. Riles* [hereafter *Riles*] labeled the EMR program “dead-end,” “isolating,” “inferior,” and “stigmatizing.” Relying on the testimony of state employees or printed material from the state department of education, the court concluded California's EMR classes were “designed to separate out children who are *incapable* of learning in regular classes [*Riles*, 1979, p. 941]” and were not meant to provide remedial instruction so that children could learn the skills necessary for eventual return to regular instruction. Given these characteristics, the court considered “the decision to place children in these classes . . . a crucial one. Children wrongly placed in these classes are unlikely to escape as they inevitably lag farther and farther behind the children in regular classes [p. 942].” And, as in *Hobson*, the primary basis for these decisions were found to be tests—most often the WISC-R and the Stanford–Binet.

Interpreting the nondiscriminatory provisions of the Rehabilitation Act of 1973 and the Education for All Handicapped Children Act of 1975 (Pub. L. 94-142), particularly regulations requiring that assessment instruments be “validated for the specific purpose for which they are used [35 C.F.R. §104.35; 34 C.F.R. §300.532],” Judge Peckham found the challenged tests unable to meet that requirement. The tests, the court ruled, would have to be shown valid for selecting children who belonged in substandard, segregated educational anachronisms (otherwise known as EMR classes). And because that kind of validation had not been done, the court permanently prohibited California “from utilizing, permitting the use of, or approving the use of any standardized tests . . . for the identification of black EMR children or their placement into EMR classes [p.

989]” without first securing the court’s approval. Even Judge Grady, who in *PASE v. Hannon* (1980) upheld the use of individual intelligence tests in a similar challenge concluded that inappropriate placement in an EMR class was an educational tragedy that was likely to be totally harmful.

Employment

Similar phenomena as I have described in public schools occurred in employment settings. *Griggs v. Duke Power Co.* (1971), which introduced the concept of “job-relatedness” into the law of employment testing and created a morass not yet fully resolved, would never have been decided if the defendants had not had a history of racial discrimination. Prior to 1965, the Duke Power Co. openly discriminated on the basis of race in the hiring and assigning of employees at its Dan River plant. Blacks were employed only in the lowest level jobs and at the lowest rate of pay. In 1964 Congress passed Title VII of the Civil Rights Act prohibiting discrimination in employment on the basis of race, religion, sex, or national origin. On July 2, 1965, the date on which Title VII took effect, Duke Power decided to no longer restrict blacks to the lowest level positions. However, at the same time, it instituted a policy that, to qualify for placement in higher level positions, employees would have to achieve satisfactory scores on the Wonderlic Personnel Test, purportedly an intelligence measure, and the Bennett Mechanical Comprehension Test. Blacks challenged the tests, claiming that neither instrument was directed or intended to measure the ability to learn to perform a particular job or category of jobs. A unanimous Supreme Court faulted the company for using “broad and general testing devices [p. 433]” and reminded the defendants that although the use of tests was permissible under Title VII, they had to “fairly measure the knowledge or skills required by the particular job [p. 433].”

Similarly, accusations of purposeful discrimination in both the private and public sector has stimulated litigation in such cases as *Albemarle Paper Co. v. Moody* (1975) and *Firefighters Institute v. City of St. Louis* (1980). The result has been increasingly sophisticated challenges to professionally developed tests even in situations where purposeful discrimination is not an issue [e.g., *Guardian Association of New York City v. Civil Service Commission* (1980); *Teal v. Connecticut* (1982)]. Like educational tests, “employment tests are being subjected to a degree of governmental scrutiny that very few human contrivances could bear [Wigdor, 1982, p. 67].”

Privacy

Although discrimination has evoked judicial scrutiny of ability tests, the concern for the right to privacy has stimulated similar examination of personality tests. If stupidity has ever fostered a judicial decision concerning testing, there is no

better example than *Merriken v. Cressman* (1973). The case had its origins in 1970 when a survey, ordered by the Commission of Montgomery County, Pennsylvania and conducted by a company called Scientific Resources, revealed that many children in the county were heavily involved with drugs. Most of the children who used drugs, the study claimed, possessed some common characteristics. For example, one finding indicated that 80% of the identified drug abusers felt estranged from their families. On the basis of such data, Scientific Resources proposed that the County Drug Commission sponsor a drug prevention program, later labeled CPI, for the Critical Period of Intervention. All three of the county school districts agreed to participate in the program.

There were two phases to the study: identification and remediation. In the first phase, tests were given to eighth-grade students and their teachers so that certain students, deemed potential drug abusers, could become part of the remediation program. The teachers were asked to identify pupils who most and least fit eight descriptions of antisocial behavior (e.g., "This pupil makes unusual or inappropriate responses during normal school activity"). The student form was to be somewhat lengthier. First, students would be asked to assess their own behavior, that is, to state which of the following statements was most like themselves: (1) someone who will probably be a success in life; (2) one who gets upset when faced with a difficult school problem; (3) someone who has lots of self-confidence; and (4) a student who has more problems than other students. In the next part of the scale they would be asked questions about their relationships with their parents and the behavior of their parents (e.g., to indicate whether one or both parents "tell me how much they love me" or "make me feel unloved" or "seem to regret that I am growing up and spending more time away from home"). Finally, the students would select from their classmates those who fit certain descriptive statements similar in kind to the ones given the teachers.

The second phase of the study was intervention. When the CPI staff had analyzed all the results, they would compile a list of children who would have significant potential for becoming drug abusers. This list would then be given to the school superintendent who would organize a joint effort among guidance counselors, teachers, school psychologists, and others to provide group therapeutic experiences to which the identified students would be involuntarily assigned.

When the program was first developed the school system did not intend to obtain the affirmative consent of the parents for their children to participate. They did plan to send a letter home to each parent, as follows:

Dear Parent:

This letter is to inform you that, this fall, we are initiating a Drug Program called "Critical Period of Intervention" (CPI). The aim of this program is to identify children who may be susceptible to drug abuse and to intervene with concrete measures to help these children. Diagnostic testing will be part of this

program and will provide data enabling the prevention program to be specific and positive.

We ask your support and cooperation in this program and assure you of the confidentiality of these studies. If you wish to examine or receive further information regarding the program, please feel free to contact the school. If you do not wish to participate in this program, please notify your principal of this decision. We will assume your cooperation unless otherwise notified by you [*Merriken v. Cressman*, 1973, p. 917].

Also, as originally proposed, the study contained no provision for student consent.

Sylvia Merriken, the mother of one of the intended participants in the study, who happened to be a therapist in a drug and alcoholic rehabilitation center, complained to the principal of the school where her son was enrolled and to the school board. The American Civil Liberties Union (ACLU) then announced it would represent Mrs. Merriken in an attempt to enjoin the school permanently from carrying out its plans. The ACLU began by filing a complaint in federal district court claiming that the program would violate the constitutional rights of both Mrs. Merriken and her son. It quickly obtained a temporary injunction prohibiting the county from implementing its proposal until the litigation was completed. At that point, two of the three schools in Montgomery County decided to discontinue their participation but the Norristown system, where Mrs. Merriken's son attended, persisted, although it honored the temporary injunction.

When the suit itself began, the school system offered to change the format of their letter to include parental consent. In another attempt at compromise, the school modified the test so that students who did not want to be included could return an uncompleted protocol. But the proposal contained no provision for student consent and no data were to be provided whereby students could make an informed choice about participating.

Of the many constitutional challenges Mrs. Merriken made, the court entertained only one of them seriously—the right of privacy. The court found that the highly personal nature of the instrument disrupted family associations and interfered with the right of the mother to rear her child. It said, “There is probably no more private a relationship, excepting marriage, which the Constitution safeguards than that between parent and child. This Court can look upon any invasion of that relationship as a direct violation of one’s Constitutional right to privacy [p. 918].” And although there was no precedent to the effect in the Supreme Court, the district court declared that privacy was entitled to as much constitutional protection as free speech.

Although the court failed to analyze the privacy rights of her son (but see Bersoff, 1983), the court found that Mrs. Merriken was unable to give genuinely informed consent to the invasion of her personal life because the parental permis-

sion letter was so inadequate. The court deridingly compared the letter to a Book-of-the-Month Club solicitation in which parents' silence would be construed as acquiescence. The letter was also criticized as a selling device in which parents were convinced to allow children to participate. It was not, as it properly should be, an objective document telling parents of the potentially negative feature and dangerous aspects of the program.

There were other problems with the program. The promotional letter promised confidentiality, but the program contemplated the development of a "massive data bank" and the dissemination of data relating to specific, identifiable students to school superintendents, principals, guidance counselors, coaches, social workers, PTA members, and school board members. And even if the school system had been more circumspect and had constructed means by which the data were less widely distributed (or not distributed at all), no promise of confidentiality could take precedence over a subpoena compelling the disclosure of the material to law enforcement officers. As the court warned:

[T]here is no assurance that should an enterprising district attorney convene a special grand jury to investigate the drug program in Montgomery County, the records of the CPI Program would remain inviolate from subpoenas and that he could not determine the identity of children who have been labeled by the CPI Program as potential drug abusers [p. 916].

Parents were not at all informed of this possibility.

Compounding the other problems was the fact that the identification instruments did not possess enough psychometric soundness to overcome the hazards that may have flowed from their use. Although there could have been considerable harm done to children correctly identified, the court was particularly concerned about those children incorrectly identified. In a statement that should raise the anxiety level of psychologists, it said, "When a program talks about labeling someone as a particular type and such a label could remain with him for the remainder of his life, the margin of error must be almost nil [p. 920]."

Forensics

Ironically, the one use of tests that has remained relatively uncriticized is in forensic assessment. Ability tests used in educational and employment decision making, despite their myriad problems, have been subjected to decades of empirical analysis and validation. Yet, they have undergone the most scathing review by the legal system. On the other hand, based on personality and projective instruments, forensic psychologists since the 1940s have routinely testified in cases involving competency to stand trial, insanity, civil commitment, the causal connection between negligent conduct and emotional and physical injury, child custody, and the eligibility of criminal defendants for the death penalty.

Whereas such determinations are at least as crucial to the interests of the test taker and society, personality and projective instruments have escaped wholesale scrutiny by the courts and remain largely untouched.

As long as psychologists possess the requisite indices of expertise such as proper education, training, experience, scholarly publications, and professional affiliation, they are permitted to offer opinions on the kinds of ultimate issues I have just cataloged based on the administration of tests like the Rorschach, MMPI, and TAT. Although such testimony has subjected individual psychologists to harsh cross-examinations (Ziskin, 1981), the courts have never seriously questioned whether these tests are sufficiently precise to evoke probative expert testimony or to support valid opinions that will be more helpful than testimony of the thoughtful layperson to the jury.

The confused approach to expert testimony by psychologists is, in part, explained by a failure to recognize that it is not a unidimensional concept but, rather, involves three levels of inference. The first level consists of the psychologist's personal observation of the client made during the course of the clinical evaluation, including essentially objective data about the individual's behavior and the uninterpreted results of psychological testing. The second level moves from reporting observations to the synthesis of data to form a diagnosis that will classify, and perhaps account for, the behavior manifested during the course of the evaluation or at the time of the event in question. It is on this level that psychologists make a judgment about whether the person has a mental disorder. Whether the diagnosis is presented in terms of a particular label or a lengthy description of personality, the critical element is that the diagnosis derives its value from the psychometric soundness of the assessment devices used. The third level concerns an opinion about the ultimate issue (i.e., child custody, or insanity) that the jury or judge must resolve. Whereas I have significant reservations about any testimony by experts as to level three (Comment, 1978; Gass & Bersoff, submitted for publication), it is concern about testimony at level two that is most relevant to this chapter.

The test for the admissibility of scientific evidence was developed 60 years ago in *Frye v. United States* (1923), which limited such evidence to that which has gained "general scientific acceptance." Under its modern interpretation by subsequent tribunals, the courts require not only acceptance within the scientific community but also accuracy. Thus, assessment devices used in the forensic arena should not only have gained acceptance within psychology but, more importantly, the accuracy of the technique should be demonstrated to yield information that is more likely to be true than could be gleaned from lay testimony. Results from polygraphs and voice spectrography have been denied admissibility because the error rate is considered to be 10 to 25%. If such a criterion were applied to most clinicians' favorite projective devices, none of the information or diagnostic conclusions derived from them would be admitted in court. Validity coefficients that clinicians might find highly acceptable may not pass

legal muster. Tests' vulnerability to situational and experimenter effects and to such phenomena as illusory correlations (Chapman & Chapman, 1969) have been well documented (Comment, 1978; Monahan, 1981).

The scientific literature regarding reliability and validity of tests used by forensic experts suggests that, at best, they are highly suspect and susceptible to a variety of significant sources of psychometric and interpretive error. They have limited psychometric soundness even in the hands of the most skilled clinicians, and there is little basis to assert that expert opinions, based on projective tests, are more accurate than layperson's opinions. But, although forensic psychologists may have little empirically based expertise to offer the legal system, they are uniformly permitted to testify and their judgments often carry great weight with the jury. On the other hand, the work of educational and industrial psychologists undergo close review, even though there is greater reason to believe that the instruments upon which they rely are more demonstrably accurate. Why? The answer is partly rooted in tradition—such issues as insanity have confronted the courts for decades; issues such as proper placement in special education programs or promotion to fire captains have not. But, more importantly, the tests used in forensic settings do not impinge on privacy or disproportionately affect racial or ethnic minorities. It is interesting to note, however, that recently forensic examinations used in criminal settings have raised concerns about the privilege against self-incrimination and right to counsel (see *Estelle v. Smith*, 1981, and *Battie v. Estelle*, 1981).

SOCIAL SCIENCE EVIDENCE AND THE COURTS

All of these conflicting perceptions about psychological tests raise a final issue with which I wish to conclude—that of the relationship between the social sciences and the courts. If that relationship were to be examined by a psychoanalyst, the analyst would no doubt conclude that it is a highly neurotic, conflict-ridden ambivalent affair (I stress “affair” because it is certainly no marriage). Thirty years ago the vitality of data generated by psychologists seemed assured when the Supreme Court in *Brown v. Board of Education* (1954) conspicuously referred to studies by Kenneth and Mamie Clark concerning the effect of segregation on black children. The reference to those studies in a now famous footnote created a controversy that still exists concerning their relevance and validity (Cahn, 1955; Clark, 1980; Kluger, 1975; Levin, 1978). Despite that controversy there is little doubt that *Brown* represents the most dramatic use of social science scholarship.

But if *Brown* produced optimism, subsequent events did not uniformly reinforce those buoyant feelings. In the past 5 years the Supreme Court has rejected empirical data in cases concerning sex discrimination, the death penalty, and corporal punishment. Perhaps most clearly exemplary of the Court's am-

bivalence is its decision in *Ballew v. Georgia* (1978) where it unanimously agreed that criminal trials before five-member juries unconstitutionally deprived defendants of the right to trial by jury. Justice Blackmun announced the judgment of the Supreme Court and in his decision relied heavily on the work of social psychologists and others to support the conclusion that less than six-person panels substantially and negatively altered the jury process. However, only one other justice joined that opinion. Three justices were particularly critical of his use of social science data. In a concurring opinion (indicating agreement with the outcome but not the reasoning of the primary opinion) Justice Powell, joined by Chief Justice Burger and Justice Rehnquist, acerbically noted his “reservations as to the wisdom . . . of Mr. Justice Blackmun’s heavy reliance on numerology derived from statistical studies [p. 246].”

The same love–hate relationship finds its way into lower-court opinions concerning testing. These opinions, regardless of whether one likes the result, are generally devoid of sound psychometric reasoning. Even if the conclusions are correct, the courts often fail to cite the relevant literature in a way that convinces the reader that the conclusion is empirically supportable.

Social Science in Education and Employment Cases

Education. We can once again return to the education and employment testing cases for the most pertinent examples. As you may recall, the court in *Riles* permanently prohibited the state from using any standardized intelligence tests for the identification of black children for placement into EMR classes and held that before the state could use IQ tests, it would have to meet the following standards:

1. Tests would have to yield the same pattern of scores when administered to different groups of students.
2. Tests would have to yield approximately equal means for all subgroups included in the standardization sample.
3. Tests would have to be correlated with relevant criterion measures, that is, IQ scores of black children with classroom performance.

The implication in *Riles* that an unbiased test must yield the same pattern of scores when administered to different groups of people is psychometrically unsound. It is generally, though not uniformly, conceded that tests are fair when they predict with equal accuracy, not with equal results, for all groups. If that position is correct, then the court’s definition “eliminates a priori any possibility of real group differences on various psychological traits [Schmidt & Hunter, 1974, p. 1].” The court rejected the possibility of genuine inferiority and social-class differences. Though the court rested its decision on the finding that the tests were culturally biased, it provided little hard data to support such a conclusion

and was tentative in discussing it. In fact, the court's empirical support for its conclusions consumed only 1 of 70 printed pages. Moreover, the court's determination that the tests contain questions biased against poor black children is not uniformly accepted, and there are some data to suggest that whatever discrimination there is in tests, lower scores in blacks are not totally the result of content bias.

By definition, achievement and intelligence tests will always fail to meet the demand for assessment devices devoid of environmental influence. Given what they purport to measure, they inevitably reflect the social setting of the test taker: "[All] behavior is . . . affected by the cultural milieu in which the individual is reared and since psychological tests are but samples of behavior, cultural influences will and should be reflected in test performance. It is therefore futile to try to devise a test that is *free* from cultural influences [Anastasi, 1976, p. 345]."

Efforts to produce culture-free tests or to reduce content bias have met with little success. "Nonverbal or performance tests are now generally recognized as falling short of the goal of freedom from cultural influences, and attempts to develop culture fair verbal tests . . . are recognized as failures [Reschly, 1979, p. 231]." More specifically, Anastasi (1976) states: "On the WISC, for instance, black children usually find the Performance Tests as difficult or more difficult than the Verbal Tests; this pattern is also characteristic of children from low socioeconomic levels [p. 348]." Kirp (1973) concludes: "[It] is sobering but instructive to recognize that minority children do poorly even on so-called culture-free tests [p. 758]."

There has been relatively little research on content bias itself, particularly with regard to individual intelligence tests. What has been found with regard to standardized tests generally (Flaughner, 1978; Green, 1978), or individual intelligence tests specifically (Reynolds, 1982; Reschly, 1980; Sandoval, 1979), does not support Judge Peckham's conclusions. For example, contrary to popular thought, such widely criticized questions on the WISC-R comprehension subtest as, "What is the thing to do if a boy (girl) much smaller than yourself starts a fight with you?" (a question that even Judge Grady in *PASE* found biased) may actually be easier for black children than they are for white (Reschly, 1979). Eliminating 13 items perceived to be biased from a widely used 82-item elementary reading test "did not improve the performance of schools with high-minority populations relative to their performance on the original 'biased version' [Flaughner, 1978, p. 675]." Deleting what appear to be idiosyncratic items from group ability tests results only in "making the tests considerably more difficult for everyone, since many of the items [exhibiting] the widest discrepancy between groups [are] moderate to low in overall difficulty [Flaughner, 1978, p. 675]" (but see Oakland & Matuszek, 1977). Most pertinently, Sandoval (1979) found no evidence of items bias on the WISC-R: "The notion that there may be a number of items with radically different difficulties for children from different ethnic groups has not been supported [p. 925]." Moreover, the interjudge agree-

ment concerning cultural bias on the WISC-R appears very low (see Reschly citing Sandoval, 1980).

Although Judge Peckham can be faulted for his analysis of cultural discrimination in intelligence tests and for implying that the issue is more settled than it is, any criticism of his analysis does not imply that his conclusion is incorrect or that there is support for such alternative hypotheses as genetics—rejected by all parties in *Riles* and *PASE*—or socioeconomic explanations. In any event, the court in *Riles* was correct in criticizing test publishers for not adequately standardizing and validating their instruments on discrete minority populations. The court could only rest its holding on the data presented to it by the parties. The state's defense was made difficult by the lack of relevant studies on differential validity, the absence of systematic research concerning content bias, and California's concession that cultural differences affected IQ scores.

If Judge Peckham's analysis of the issue of cultural bias was scanty and faulty, Judge Grady in *PASE v. Hannon* (1980) can best be described as naive. At worst it was unintelligent, and completely devoid of empirical content. Distrustful of the expert testimony in the case, he felt it imperative to examine the tests themselves so he could judge whether the claim of cultural bias could be sustained. Thus, in a startling and extraordinary manner, he proceeded to cite each question on the Wechsler and Binet scales in an attempt to determine which, in his estimation, were culturally biased. The result of this analysis was the judgment that only eight items on the WISC/WISC-R and one item on the Stanford-Binet were suspect or actually biased. At bottom, what it represented was a single person's subjective and personal judgment cloaked in the apparent authority of judicial robes. If submitted as a study to one of psychology's more respected refereed journals, rather than masquerading as a legal opinion, it would have been summarily rejected as an experiment whose sample size and lack of objectivity stamped it as unworthy of publication. The court's opinion in *PASE* amply supports Reschly's (1980) conclusion that with regard to item bias on the individually administered intelligence tests, "subjective judgments appear to be unreliable and invalid in terms of empirical analysis. . . . The only data confirming test bias that exists now is judgmental and speculative [p. 127]."

What makes Judge Grady's opinion interesting, if not precedent setting, is the fact the decision contains the questions and correct answers to every item on the WISC, the WISC-R, and the Stanford-Binet. McClelland (1973) suggested several years ago that tests should be given away. Whether inadvertently or purposely, Judge Grady has done just that. Those who wish to destroy the usefulness of these tests need only inform parents and antitest advocates of the existence of the decision and its citation to the proper volume in the series of legal reports that publishes verbatim all federal district court opinions. Although Judge Grady eventually upheld the tests as valid, his decision, to a far greater extent than Judge Peckham's decision in *Riles*, may have the effect of invalidating the tests as they are presently used. The Psychological Corporation, publisher

of the Wechsler Scales (and the System of Multi-Pluralistic Assessment [SOMPA] that uses these scales), tried unsuccessfully to convince Judge Grady to seal that part of his decision containing the questions and answers to the scales so that their content would not be published and thus made public. It has since issued a statement attempting to protect its copyright in the tests and threatens legal action if it is not protected: "The Psychological Corporation considers unauthorized reproduction of its copyrighted material from any source, including a court's opinion, to be an invasion of its rights, including its copyright, and the right to maintain the necessary security of its tests [Udell, 1980]." As of this writing, there has not been specific legal action against those who have informed general audiences of its existence. But one potential outcome of the decision is that the security of these tests may have, indeed, been seriously compromised, if not destroyed.

Employment. The situation with regard to employment testing does not evoke any greater confidence. There are sharp differences among the federal courts, Equal Employment Opportunity Commission (EEOC), and psychometric experts as to the proper conceptualization of test validation within the industrial setting. Novick (1981) has perceptively summarized the struggle:

Individual federal agencies have responsibilities and goals delegated by the executive and legislative branches of government, monitored by the judicial branch, and ultimately specified by the incumbent agency management. Although these agencies share concern for benefits to society as a whole, they tend to focus attention on their own particular mandates, and for this reason they often view testing and other issues quite differently. In fact, it is not uncommon for government agencies to be on opposite sides in litigation involving tests, for employers to receive conflicting directives from different government agencies, and for employees to find that their test scores are considered in light of widely varying objectives by employers and government agency representatives [p. 1035].

The Supreme Court has been particularly unhelpful in sorting out this confusion. For example, in *Griggs v. Duke Power Co.* (1971) a unanimous court stated that the EEOC Guidelines on Employment Testing were "entitled to great deference" (p. 434). Four years later, Chief Justice Burger, who had written the decision in *Griggs* now complained in a minority opinion in *Albermarle Paper Co. v. Moody* (1975) about the Court's "slavish adherence" (p. 452) to those same Guidelines. Perhaps in a more important example, one I described at some length in a recent *American Psychologist* article (Bersoff, 1981), the Court has badly muddled the whole issue of test validation. In *Washington v. Davis* (1976) in support of its opinion that validation could be accomplished in "any one of several ways," the Court cited the then extant version of the *Standards for Educational and Psychological Tests* (APA, AERA, NCME, 1974) to the effect

that there were “three basic methods of validation: ‘empirical’ or ‘criterion’ validity . . . ‘construct validity’ . . . and ‘content’ validity [*Washington v. Davis*, 1976, p. 247, fn. 13].”

Many industrial and academic psychologists (Guion, 1980; Messick, 1980; Tenopyr, 1977) contend that insofar as the courts have interpreted the test standards and the EEOC Guidelines (superseded now by the Uniform Guidelines on Employee Selection Procedures, see EEOC et al., 1978) and its implementing “Question and Answers” (EEOC et al., 1979) to mean that content, criterion, and construct validity are distinct forms of validation, those interpretations are oversimplified, if not erroneous. The Uniform Guidelines, according to this view, inappropriately treat three aspects of validity as “something of a holy trinity representing three different roads to psychometric salvation [Guion, 1980, p. 386]” rather than viewing them as subsets within the unifying and common framework of construct validity. Most judicial opinions, with one or two conspicuous exceptions (see *Guardians Association of New York City v. Civil Service Commission*, 1980), concerned with the controversy over content versus criterion versus construct validity in employment tests also view the three as separable entities rather than on a continuum and fail to cite or even recognize the work of psychologists who have urged a more sophisticated approach to validation analysis. It has been suggested that the term *construct-referenced* validity (Messick, 1975) would more precisely encompass almost all discrete and specialized validation terms, integrating content relevance and content coverage as well as predictive and diagnostic utility. “The bridge or unifying theme that permits this integration is the meaningfulness of interpretability of the test scores, which is the goal of the construct validation process [Messick, 1980, p. 1015].”

In 1982, the Supreme Court had the opportunity to review the EEOC Guidelines and its implementing Questions and Answers in *Teal v. Connecticut* (1982) but carefully avoided the issue. In that case, the plaintiffs are four black provisional state employees who, when they sought to attain permanent status in their jobs as Welfare Eligibility Supervisors, were obliged to participate in a selection process requiring a passing score on a written test. Those who passed the test became part of an eligibility pool from which the state would select successful applicants. The final determinations were made on the basis of a number of nontest criteria (e.g., past work, recommendation).

All the plaintiffs failed to achieve the cutoff score of 65 on the test which would have made them eligible for further consideration. As a whole, the passing rate for blacks was 68% of that of whites. The unsuccessful plaintiffs then instituted a suit claiming that the state’s use of the test violated Title VII. However, a month prior to trial, the state made its final selection, the result of which was that 23% of the eligible blacks and 13.5% of the eligible whites were promoted to supervisor. The actual promotion rate of blacks, therefore, was 169.5% of the actual promotion rate of whites. Thus, whereas the end result of

the state's selection process (the so-called "bottom line") was nondiscriminatory to blacks as a class, the threshold testing component did not meet the Uniform Guidelines "four-fifths" rule, which provides that a "selection rate for any race, sex, or ethnic group which is less than [80%] of the rate for the group with the highest rate will generally be regarded . . . as evidence of adverse impact [29 C.F.R. §1607.4(c)]."

The federal district court dismissed the plaintiffs' claims, holding that they failed to prove a *prima facie* case of disparate impact. It asserted that although the ratio of the black *passing* rate to the white passing rate was 68%, the ratio of the black *appointment* rate to the white appointment rate was almost 170%. Thus, under the bottom-line approach found in the EEOC Guidelines, the plaintiffs' Title VII claim has to fail.

The plaintiffs appealed. The Court of Appeals reversed the lower court, holding that "where a plaintiff establishes that a component of a selection process produced disparate results *and* constituted a pass-fail barrier beyond which the complaining candidates were not permitted to proceed, a *prima facie* case of disparate impact is established, notwithstanding that the entire selection procedure did not yield disparate results [*Teal v. Connecticut*, 1981, p. 135]."

In concluding that the district court was wrong in ruling results of the written examination alone were insufficient to support a *prima facie* case of disparate impact, it distinguished an earlier decision by the second circuit court. In *Kirkland v. New York State Dept. of Correctional Services* (1975), the Court of Appeals held that proof concerning disparate impact of certain subtests within a larger examination did not constitute an unlawful discriminatory impact. But, the second circuit said in *Kirkland*, all applicants were subjected to a complete selection process that, when viewed as a whole, did not produce disparate results. In *Teal*, however, the pass-fail barrier denied employment opportunity to a disproportionately large number of minorities and prevented them from proceeding to the next step in the selection process. Thus, the court concluded, affirmative action policies that may benefit minority groups as a class do not excuse employers' discriminatory conduct affecting specific and readily identifiable individuals. It held that "Title VII was designed to protect the rights of individuals" and that it "matters very little to the victimized individuals that their group as a whole is well represented in the group of hires [pp. 139-140]."

The trial court, finding no evidence of *prima facie* discrimination, never reached the question of the test's validity (i.e., its "job-relatedness"), even though it had been fully tried before the court. However, in addition to reversing the trial court's decision, the Court of Appeals remanded the case with instructions that the lower court evaluate the test itself in light of the EEOC Guidelines.

The state of Connecticut, in June of 1981, asked the Supreme Court to review the second circuit's opinion, arguing that their decision was antagonistic to that of other circuits who had adopted the bottom-line concept in Title VII cases. The state also asserted that scrutiny of testing practices in those instances where

hiring or promotion practices revealed no disparate impact would redirect employers' concerns from "the overall hiring process to the testing process, and in that sense [the federal courts would] be restructuring business practices."

The Supreme Court agreed to review the case and in June 1982 it rendered its opinion. The Court held, in a 5–4 decision, that "the 'bottom line' does not preclude . . . employees from establishing a prima facie case [of employment discrimination] nor does it provide [an] employer with a defense to such a case [*Teal v. Connecticut*, 1982, p. 2529]." The Court reminded employers that Section 703(a)(2) spoke not in terms of jobs and promotions but of limitations and classifications that would deprive individuals of employment opportunities. Thus, "when an employer uses a non-job-related barrier to deny a minority or woman applicant employment or promotion, and that barrier has a significant adverse effect on minorities or women, then the applicant has been deprived of an employment *opportunity* 'because of . . . race, color, religion, sex, or national origin' [p. 2532]." Therefore, Title VII protects individuals, not groups prohibiting victims of a facially discriminatory policy to be told that they have not been wronged simply because other persons of their race or sex were hired: "Every *individual* employee is protected against both discriminatory treatment and against practices that are fair in form, but discriminatory in operation. . . ." [p. 2535] [emphasis added]."

As a result, the Court refused to permit employers to claim as a defense in disparate impact cases that discriminatory, non-job-related tests that serve as a pass–fail barrier to employment opportunities are permissible because the tests did not actually deprive disproportionate numbers of blacks of promotions. "It is clear," the Court asserted, "that Congress never intended to give an employer license to discriminate against some employees on the bias of race or sex merely because he favorably treats other members of the employees' groups [p. 2535]."

The dissenters, led by Justice Powell, speaking for the Chief Justice and Justices Rehnquist and O'Connor, agreed that the aim of Title VII was to protect individuals, not groups. But, they interpreted disparate impact claims to require proof of discrimination to groups. The dissenting opinion argued that prior cases had made it clear that discriminatory impact claims cannot be based on how an individual is treated because those claims are necessarily based on whether the group fares less well than other groups under a policy, practice, or test. The dissent warned that the majority's holding could "force employers either to eliminate tests or rely on expensive, job-related, testing procedures, the validity of which may or may not be sustained if challenged. For state and local governmental employers with limited funds, the practical effect of today's decision may well be the adoption of simple quota hiring [p. 2540]." Moreover, it cautioned, substantially fewer minority candidates ultimately could be hired simply by employers integrating consideration of test results into one overall hiring decision because, by so doing, "they will be free to select *only* the number of

minority candidates proportional to their representation in the workforce [p. 2540 n.8].”

All these decisions reveal that the issue of test bias is complex and controversial and that opinions concerning its existence are contradictory. Several models of test bias, particularly with regard to its effect on prediction and selection, have been offered (Jensen, 1980; Peterson & Novick, 1976), none of which seem to have gained favor over others. As Ysseldyke (1978) recently commented:

Several investigators have reviewed the models of test fairness and have concluded that there is little agreement among the several models. It is readily apparent that major measurement experts have been essentially *unable* to agree on a definition of a fair test, let alone identify a test that is fair for members of different groups. There is little agreement on the *concept* of nondiscriminatory assessment [p. 150].

Definitions of test bias may not only be “widely disparate,” stemming “from entirely different universes of discourse [Schmidt & Hunter, 1974, p. 1]” but ethical positions regarding test bias may be “irreconcilable [Hunter & Schmidt, 1976, p. 1069].” Finally, and perhaps more importantly, reliance on psychometric models for test bias without consideration of the social and ethical consequences of test use ignores the concerns of significant segments of society. Although the American Psychological Association Ad Hoc Committee Report on the Educational Uses of Tests with Disadvantaged Students (Cleary et al., 1975) defended the technical adequacy of tests for prediction and selection, it failed to consider what minority groups charge was the egregious misuse of tests having a negative impact on the lives of minorities (Bernal, 1975; Jackson, 1975). As Reschly (1979) points out: “to defend tests on the basis of evidence of common regression systems or to attempt to separate the issues of technical adequacy from the social consequences is insufficient [p. 235].” In that light, recent attempts to examine the ethical, legal, and social implications of various models of test bias are valuable additions to the literature (Cole, 1981; Hunter & Schmidt, 1976; Messick, 1980; Novick & Ellis, 1977; Reynolds, 1982). In essence, even the selection of a model to measure and ameliorate test bias is ultimately a value judgment (Kaplan, 1982).

PSYCHOLOGISTS AND PUBLIC POLICY

My complaints about the Supreme Court should not deflect responsibility from psychologists. I think it may be legitimate to place at least part of the fault for the current and continuing confusion concerning tests on psychologists themselves. One of the more intriguing aspects of Judge Grady’s decision in *PASE v. Hannon*

(1980) was his almost utter rejection of the testimony of expert psychologists who testified either for the black children challenging the IQ tests or for the school system seeking to defend them. In a quote that I think deserves some thought he said:

None of the witnesses in this case has so impressed me with his or her credibility or expertise that I would feel secure in basing a decision simply on his or her opinion. In some instances, I am satisfied that the opinions expressed are more the result of doctrinaire commitment to a preconceived idea than they are the result of scientific inquiry. I need something more than the conclusions of witnesses in order to arrive at my own conclusion [p. 836].

Several years ago Cronbach (1975) warned psychologists involved in testing issues not to be advocates. But, far too often they have testified *for* one side or the other. Although psychologists perform a valuable service when they testify as expert witnesses, they should be aware that their data, interpretations, and opinions will be tested in the crucible of courtroom cross-examination whose very purpose is to destroy credibility and evoke evidence of bias on the part of the expert. Whereas the distillation of that process may yield testimony of great consequence and weight to the court, it can be highly anxiety provoking for the psychologist who acts as an injudicious advocate pleading for a position rather than as a cautious, neutral scientist presenting data in an even-handed manner.

Recently, concerned psychologists have indicated the many ways social scientists can influence public policy effectively (Bersoff, 1983; DeLeon, O'Keefe, Vandenbos, & Kraut, 1982; Horowitz & Katz, 1975; Loftus & Monahan, 1980; Saks, 1978). Within the bounds of scientific and professional ethics, that is an important, if not crucial, role. But, if psychologists are to be respected by the courts and treated as more than mere numerologists attempting to convince the judiciary of doctrinaire positions, they must offer more situation-specific, ecologically valid, objective data that serve science, not a particular adversary. In that way, perhaps, courts may finally arrive at not only judicially sound but psychometrically justified decisions that will withstand both appellate and scientific scrutiny.

ACKNOWLEDGMENTS

One receives invitations to prestigious conferences like this only after one has written a great deal on a particular topic. Unfortunately, by the time the invitation arrives, the writer has said almost everything he or she has to say. I am afraid this is true about this chapter. Although I have tried to be original, much of the material is based on prior publications, most especially Bersoff (1979, 1981, 1982a, 1982b).

REFERENCES

- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. *Standards for educational and psychological tests*. Washington, D.C.: American Psychological Association, 1974.
- Anastasi, A. *Psychological testing* (4th ed.). N.Y.: Macmillan, 1976.
- Bernal, E. A response to "Educational uses of tests with disadvantaged students." *American Psychologist*, 1975, 30, 93-95.
- Bersoff, D. Regarding psychologists testily: Legal regulation of psychological assessment in the public schools. *Maryland Law Review*, 1979, 39, 27-120.
- Bersoff, D. Testing and the law. *American Psychologist*, 1981, 36, 1047-1056.
- Bersoff, D. Larry P. and PASE: Judicial report cards on the validity of individual intelligence tests. In T. Kratochwill (ed.), *Advances in school psychology* (vol. 2). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1982. (a)
- Bersoff, D. The legal regulation of school psychology. In C. Reynolds & T. Gutkin (Eds.), *The handbook of school psychology*. New York: Wiley, 1982. (b)
- Bersoff, D. Children as participants in psychoeducational assessment. In G. Melton, G. Koocher, & M. Saks (Eds.), *Children's competence to consent*. New York: Plenum, 1983.
- Cahn, E. Jurisprudence. *New York University Law Review*, 1955, 30, 150-169.
- Chapman, L., & Chapman, J. Illusory correlations as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 1969, 74, 271-280.
- Clark, K. Racial justice in education: Continuing struggle in a new era. *Howard Law Journal*, 1980, 23, 93-106.
- Cleary, A., Humphreys, L., Kendrick, S., & Wesman, A. Educational uses of tests with disadvantaged students. *American Psychologist*, 1975, 30, 15-41.
- Cole, N. Bias in testing. *American Psychologist*, 1981, 36, 1067-1077.
- Comment. Psychologist as expert witness: Science in the courtroom? *Maryland Law Review*, 1978, 38, 539-621.
- Cronbach, L. Five decades of public controversy over mental testing. *American Psychologist*, 1975, 30, 1-14.
- DeLeon, P., O'Keefe, A., Vandenbos, G., & Kraut, A. How to influence public policy: A blueprint for activism. *American Psychologist*, 1982, 37, 476-485.
- Dubois, B. A test-dominated society: China, 1115 B.C.-1905 A.D. In A. Anastasi (Ed.), *Testing problems in perspective*. Washington, D.C.: American Council on Education, 1966.
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. Adoption by four agencies of Uniform Guidelines on Employee Selection Procedures. *Federal Register*, 1978, 43, 38920-39315.
- Equal Employment Opportunity Commission, Office of Personnel Management, Department of Justice, Department of Labor, & Department of the Treasury. Adoption of questions and answers to clarify and provide a common interpretation of the Uniform Guidelines on Employee Selection Procedures. *Federal Register*, 1979, 44, 11996-12009.
- Flaugher, R. The many definitions of test bias. *American Psychologist*, 1978, 33, 671-679.
- Gass, R., & Bersoff, D. *Psychologist as expert witness: Taking a stand*. (Manuscript submitted for publication).
- Green, B. In defense of measurement. *American Psychologist*, 1978, 33, 664-670.
- Guion, R. On trinitarian doctrines of validity. *Professional Psychology*, 1980, 11, 385-398.
- Haney, W. Validity, vaudeville, and values. *American Psychologist*, 1981, 36, 1021-1034.
- Horowitz, J., & Katz, J. *Social science and public policy in the United States*. New York: Praeger, 1975.

- Hunter, J., & Schmidt, F. Critical analysis of the statistical and ethical implications of various definitions of test bias. *Psychological Bulletin*, 1976, 83, 1053–1071.
- Jackson, G. On the report of the Ad Hoc Committee on Educational Tests with Disadvantaged Students: Another psychological view from the Association of Black Psychologists. *American Psychologist*, 1975, 30, 88–92.
- Jensen, A. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, 39, 1–123.
- Jensen, A. *Bias in mental testing*. New York: Free Press, 1980.
- Kaplan, R. Nader's raid on the testing industry. *American Psychologist*, 1982, 37, 15–23.
- Kirp, D. Schools as sorters: The constitutional and policy implications of student classification. *University of Pennsylvania Law Review*, 1973, 121, 705–797.
- Kluger, R. *Simple justice: The history of Brown v. Board of Education and black America's struggle for equality*. New York: Knopf, 1976.
- Levin, B. School desegregation remedies and the role of social science research. *Law and Contemporary Problems*, 1978, 42, 1–36.
- Loftus, E., & Monahan, J. Trial by data: Psychological research as legal evidence. *American Psychologist*, 1980, 35, 270–283.
- McClelland, D. Testing for competence rather than for intelligence. *American Psychologist*, 1973, 28, 1–14.
- Messick, S. The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 1975, 30, 955–966.
- Messick, S. Test validity and the ethics of assessment. *American Psychologist*, 1980, 35, 1012–1027.
- Monahan, J. *The clinical prediction of violent behavior*. Washington, D.C.: NIMH, 1981.
- Novick, M. Federal guidelines and professional standards. *American Psychologist*, 1981, 36, 1035–1046.
- Novick, M., & Ellis, D. Equal opportunity in educational employment selection. *American Psychologist*, 1977, 32, 306–320.
- Oakland, T., & Matuszek, P. Using tests in nondiscriminatory assessment. In T. Oakland (Ed.), *Psychological and educational assessment of minority children*. New York: Brunner/Mazel, 1977.
- Peterson, N., & Novick, M. An evaluation of some models for culture-fair selection. *Journal of Education Measurement*, 1976, 13, 3–29.
- Reschly, D. Nonbiased assessment. In G. Phye & D. Reschly (Eds.), *School psychology: Perspective and issues*. New York: Academic Press, 1979.
- Reschly, D. Psychological evidence in the Larry P. opinion: A case of right problem—wrong solution? *School Psychology Review*, 1980, 9, 123–135.
- Reubhausen, D., & Brim, O. Privacy and behavioral research. *Columbia Law Review*, 1965, 65, 1184–1215.
- Reynolds, C. The problem of bias in psychological assessment. In C. Reynolds & T. Gutkin (Eds.), *The handbook of school psychology*. New York: Wiley, 1982.
- Saks, M. Social psychological contributions to a legislative subcommittee on organ and tissue transplants. *American Psychologist*, 1978, 33, 690–690.
- Sandoval, J. The WISC-R and internal evidence of test bias and minority groups. *Journal of Consulting and Clinical Psychology*, 1979, 47, 919–927.
- Schmidt, F., & Hunter, J. Racial and ethnic bias in psychological tests: Divergent implications of two definitions of test bias. *American Psychologist*, 1974, 29, 1–8.
- Tenopyr, M. Content-construct confusion. *Personnel Psychologist*, 1977, 30, 47–54.
- Udell, R. Personal communication with *APA Monitor*, Nov. 1, 1980.

- Wigdor, A. Psychological testing and the law of employment discrimination. In A. Wigdor & W. Garner (Eds.), *Ability testing: Uses, consequences, and controversies*. Washington, D.C.: National Academy of Sciences, 1982.
- Ysseldyke, J. Implementing the "Protection in Evaluation Procedures" provisions of Public Law 94-142. In L. Morra (Ed.), *Developing criteria for the evaluation of protection in evaluation procedures provisions*. Washington, D.C.: Bureau of Education of the Handicapped, 1978.
- Ziskin, J. *Coping with psychiatric and psychological testimony*. (3d ed.). Beverly Hills, Calif.: Law and Psychology Press, 1981.

