

EXHIBIT 52



**SOCIETY for
INDUSTRIAL and
ORGANIZATIONAL
PSYCHOLOGY**

Home | /

SCIENCE FOR A SMARTER WORKPLACE

Division 14 of the American Psychological Association
(APA) and Organization for Psychological Science

info for: ▶

PROFESSIONALS

STUDENTS

EDUCATORS

MEDIA

MEMBERSHIP

MEETINGS

SERVICES

PUBLICATIONS

JOBS

RESOURCES

FOUNDATION

PARTNERS

my.SIOP

Access your profile, account preferences, member directory, and member-only services

Access



What is I-O?

Industrial-organizational (I-O) psychology is the scientific study of the workplace. Rigor and methods of psychology are applied to issues of critical relevance to business, including talent management, coaching, assessment, selection, training, organizational development, performance, and work-life balance.

Find an I-O Job

I-O Graduate Programs

My.SIOP Community

SIOP Social Media



OCR Issues Draft Guide on Disparate Impact in Educational Testing

**Wayne Camara
The College Board**

In May, the Department of Education's Office of Civil Rights (OCR) released a draft Resource Discrimination in High Stakes Testing" that sought to provide an overview of federal standards principles that should guide the use of tests for making high stakes educational decisions (e.g. special educational referrals, promotion, graduation, and scholarship awards). This Resource development for several years according to OCR, but educational groups and test publishers working days before it was originally scheduled for release.

The Guide may have limited direct impact on I-O psychologists, unless they are involved in education. However, the Guide may be of interest for other reasons, since it interprets and applies both I-O employment arena and professional testing standards to issues of disparate impact in ways that "overreaching" or incorrect.

Test publishers, APA, and other educational institutions objected to the proposed timing of the Guide. OCR has agreed to revise the current document with plans for a fall publication. OCR has stated the Guide will provide new federal guidelines or professional standards, but rather will provide a meaningful interpretation of existing tests in education. A number of national media outlets (*New York Times*, *Wall Street Journal*, *Chronicle of Higher Education*) have run stories on the guidelines and op-ed pieces that have emphasized on disparate impact being the sole determination of whether or not a test should be used.

The Guide cites specific wording from the *Standards for Educational and Psychological Testing* on 40 occasions, leading APA, AERA, and NCME to formally request that OCR delay revision of the Guide until it has been revised and published (sometime around December 1999). Several organizations have submitted comments on the OCR Guide.

The Guide attempts to apply Title VII law, EEOC Guidelines, and professional standards that apply to educational test use. It cites several Supreme Court and lower courts decisions concerning employment or transports decisions and standards to education. Major concerns addressed by educationists are summarized in comments submitted by the College Board (Camara, 6/21/1999):

First, the Resource Guide focuses exclusively on disparate impact resulting from tests (or differences) and ignores the level of validity and utility offered by a test. Disparate impact cannot be considered in isolation; it must be evaluated in terms of the overall validity and utility of inferences associated with the test. The Resource Guide clearly elevates any measure, irrespective of validity, cost, or burden to the test, if it results in lower disparate outcomes above any test having greater disparate outcomes. We believe this is a bad precedent that has no legal or professional justification and the Guide will have a chilling effect on the use of educational tests.

Second, the Resource Guide offers no guidance on what level of disparate impact would result in an investigation. There should be substantial statistical disparities or would any disparate outcome result in an investigation?

should not be the primary statistical analysis used to determine if and when an alternative measure is appropriate. A consistent pattern of ethnic and racial disparities has been found across a variety of standardized tests, including the National Assessment of Educational Progress (NAEP) and the National Educational Longitudinal Study. Educational measures used for high-stakes decisions, such as high school grades, class rank, and quality and rigor of courses completed, as well as educational outcomes (e.g., college grades) (Camara and Schmidt, under review). Disparities in test results reflect similar differences in other measures (e.g., job performance, college achievement, and grades) and may be indicative of earlier differences in learning and educational opportunities, not test bias or flaws with the test.

Third, professional and technical standards do not define tests so narrowly that they exclude other measures. Assessments that are both used daily to make high-stakes decisions about individual students and have similar levels of disparate impact against protected groups. Specifically, the Test Standards include standardized ability (aptitude and achievement) instruments, diagnostic and evaluative devices, personality inventories, and projective instruments. A more appropriate choice among assessment measures will be facilitated if there is a reasonable comparability in the kinds of information available. The three broad categories of test instruments **are covered** [emphasis added]: constructed performance tests and to a lesser extent, structured behavioral samples (pages, 3_4)." Related to this comment, the Resource Guide be renamed to put added emphasis on Measures Used in Making High-Stakes Decisions (*Uniform Guidelines and Employment Selection Procedures*), rather than focus exclusively on the decision-making process, testing.

Fourth, we applaud OCR's deference to the Test Standards. However, the Resource Guide in its application of professional standards can be applied in a rigid manner in evaluating tests. The Test Standards are not a rigid checklist approach, noting that specific circumstances affect the relevance of standards. Standards must be applied in evaluating tests. Professional practice and standards are typically constructed based on other measures need not meet all standards to be appropriately used within the bounds of professional practice (Richardson, 729 F. Supp. At 821, 823). In addition, the three sponsoring educational associations have endorsed the Standards, which date back to 1985. We strongly endorse the recommendations from APAC asking that issuance of this Resource Guide be deferred until after publication and dissemination of the Standards and requesting a standard 90-day review period for any subsequent drafts of this Resource Guide prior to publication of the revised Test Standards.

Fifth, we would ask OCR to ensure that colleges and universities, school districts, and state education agencies have an opportunity to review and comment on this proposed Resource Guide. The Resource Guide should be disseminated or reviewed by colleges and secondary schools. These are the very organizations most affected by the Resource Guide once it is issued and it seems appropriate that they be given an opportunity to comment on the inferences and proposed standards.

Sixth, the distinction the Resource Guide makes between tests and other assessment devices is not supported by establishing a much lower technical, professional, and legal standard for more subjective assessment devices such as applications, grades and GPA, recommendations, ratings or evaluations of student work and experiences and honors, community service and involvement, samples of student work). In *Watson v. Fort Worth Bank and Trust*, 487 U.S. 977, the American Psychological Association submitted an amicus curiae brief arguing there is no professional or scientific justification to treat subjective and objective devices differently. In fact, not imposing essentially the same legal and technical standards for tests and devices used in high-stakes individual decisions would provide a sanctioned and covert route for discrimination. APA further argued that subjective procedures (in that case used for employment) are "amenable to psychometric scrutiny" as objective procedures, citing the Test Standards which address integrity of assessment (Camara, 1996). In deciding *Watson v. Fort Worth Bank and Trust*, 487 U.S. 977, all eight of the Justices agreed with O'Connor's opinion holding that the adverse impact theory can be used in cases involving subjective assessment. O'Connor was concerned that an employer could combine an objective criteria (such as a test or diploma) with subjective criteria (such as interviews or ratings) and easily insulate itself from the Griggs test. O'Connor noted that such a system could have "precisely the same effects as a system pervaded by impermissible discrimination" (Opinion at 4926).

Seventh, professional and legal standards do not provide any support for OCR's distinctions between tests and other measures. We agree with comments to an earlier draft of this Resource Guide submitted by the American Psychological Association in *Testing and Assessment* (Shavelson, June 10, 1996), stating that "OCR's inquiry is not to prove the validity of inferences and decisions based on tests, but rather to determine whether the entire assessment system as a part provides students a fair and equal opportunity to learn...." The Resource Guide ignores the fact that tests and other measures even if they contribute more to disparate outcomes. In fact, high school courses, judgments a

high school curriculum," grades, and rank may also contribute more to disparate outcomes, if testing, if an institution places substantially greater weight on these factors. For example, if one of several factors in admissions, then there is no guarantee those disparate outcomes will be eliminated. In requiring tests to meet an exceptionally higher standard than other measures (e.g., student work, high school rank, past experience, and opportunities), the Resource Guide will not be of valid and objective standardized tests used by educational institutions, states, and schools. Educational institutions may opt to employ less valid and less objective methods for high-stakes tests that are not addressed in this Resource Guide.

Eighth, the Resource Guide also sanctions the use of the *Uniform Guidelines on Employment Testing* as a resource in educational testing. As the Resource Guide acknowledges in a footnote, there are significant differences between educational and employment testing that we believe undermines any attempt to apply the *Uniform Guidelines* in educational settings. The *Uniform Guidelines* were never developed with application to educational testing. Educational organizations did not have an opportunity to comment on extensions of the principles to educational testing. The *Guidelines* are over 25 years old and do not reflect current scientific principles of measurement and testing practice. The *Uniform Guidelines* are outdated and do not conform to the Testing Standards (APA, 1985) in their consideration of validity (as accomplished by adopting one of three distinct types of validity: content, construct, and criterion) (this is virtually ignored in the *Guidelines*, but is accepted professional practice), differential item functioning, as well as several other areas (APA, 1985). The *Uniform Guidelines* may provide a framework for developing guidelines addressing test use, but they should not be viewed as a substantive resource in educational testing.

Ninth, statistical analyses should be based on the pool of qualified applicants, not a general population. This issue is not addressed in the Resource Guide.

Tenth, this Resource Guide implies that once disparate impact is established that the burden shifts to the educational institution to demonstrate both the educational necessity of the test and then to demonstrate that an alternative exists throughout the process. This legal interpretation is incorrect.

Other sections of the Resource Guide viewed as problematic include wording implying that separate studies are required for each school; that tests can only be used for purposes they were originally designed for (rather than for uses where sufficient validation evidence exist); and that there is a unique methodology for each test when they are to be used as the sole criteria.

On June 18th, the House held a hearing on the OCR Guide and department officials noted that they would like to recirculate the current draft to groups who have already submitted comments on the current draft and submit a revised Guide to the National Academy of Sciences Board of Testing and Assessment. Thereafter, they anticipate making a final draft available for public review this fall. They will post the draft on their web and will have the revised Guide posted on their web (Coleman, June 21, 1999, personal correspondence).

References:

American Educational Research Association, American Psychological Association, and National Academy of Sciences Board of Testing and Assessment (1985). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychological Association (1985). *Report of the Ad Hoc Committee to Develop a Uniform Guidelines on Employee Selection Procedures*. Washington, DC: Author.

American Psychological Association (1987). *Amicus curiae brief in support of the petitioner. V. Watson v. Fort Worth Bank and Trust*. Washington, DC: Author.

Camara, W. (1996). Fairness and public policy in employment testing: Influences from a professional perspective. In W. Barrett (Ed.), *Fair employment strategies in human resource management*. Westport, CT: Quorum Books.

Camara, W. and Schmidt, A. (Under Review). Social stratification and group differences in standardized test scores and educational indicators.

Shavelson, R. (June 10, 1996). *Correspondence to Norma Cantu*.

Watson v. Fort Worth Bank and Trust, 487 U.S. 977 (U.S. Supreme Court, June 29, 1988).

