


EXHIBIT 68



[Home](#) | [About Us](#) | [Sitemap](#) | [Contact](#)

SOCIETY FOR INDUSTRIAL & ORGANIZATIONAL PSYCHOLOGY, INC.

Division 14 of the American Psychological Association (APA) and Organizational Affiliate of the Association for Psychological Science (APS)

"Integrating science and practice at work."

info for: ▶
PROFESSIONALS
STUDENTS
EDUCATORS
MEDIA

powered by Google

[MEMBERSHIP](#) | [MEETINGS](#) | [SERVICES](#) | [PUBLICATIONS](#) | [JOBS](#) | [RESOURCES](#) | [FOUNDATION](#) | [PARTNERS](#)

Member Login

Username:

Password:

[Forgot Password?](#)

LOGIN

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

Graduate Training Programs in I-O

Donate to the SIOP Foundation

SIOP Exchange Official I-O Blog

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 Guide entitled, "Non-Discrimination in High Stakes Testing" that sought to provide an overview of federal standards and related educational principles that should guide the use of tests for making high stakes educational decisions (e.g., placement, admissions, special educational referrals, promotion, graduation, and scholarship awards). This Resource Guide has been under development for several years according to OCR, but educational groups and test publishers only received a copy 5 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 educational decision making. However, the Guide may be of interest for other reasons, since it interprets and applies both legal court decisions from the employment arena and professional testing standards to issues of disparate impact in ways that many may see as "overreaching" or incorrect.

Test publishers, APA, and other educational institutions objected to the proposed timing of the release and OCR has agreed to revise the current document with plans for a fall publication. OCR has stated the Guide is not establishing any new federal guidelines or professional standards, but rather will provide a meaningful interpretative tool for those who use tests in education. A number of national media outlets (*New York Times*, *Wall Street Journal*, *US News and World Report*, *Chronicle of Higher Education*) have run stories on the guidelines and op-ed pieces that have largely been critical of the emphasis 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* (Test Standards) on over 40 occasions, leading APA, AERA, and NCME to formally request that OCR delay revision of the document until after it has been revised and published (sometime around December 1999). Several organizations have now provided detailed comments on the OCR Guide.

The Guide attempts to apply Title VII law, EEOC Guidelines, and professional standards that apply to employment testing to educational test use. It cites several Supreme Court and lower courts decisions concerning Title VII issues and applies or transports decisions and standards to education. Major concerns addressed by educational organizations were 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 in outcomes) and ignores the level of validity and utility offered by a test. Disparate impact cannot be considered in isolation, but rather it must be evaluated in terms of the overall validity and utility of inferences associated with the particular test use. The Resource Guide clearly elevates any measure, irrespective of validity, cost, or burden to the educational institution, with lower disparate outcomes above any test having greater disparate outcomes. We believe this is a dangerous new precedent that has no legal or professional justification and the Guide will have a chilling effect on institutions who use educational tests.

Second, the Resource Guide offers no guidance on what level of disparate impact would result in an investigation. Must there be substantial statistical disparities or would any disparate outcome result in an investigation? Disparate outcomes should not be the primary statistical analysis used to determine if and when an alternative measure should be employed. A consistent pattern of ethnic and racial disparities has been found across a variety of standardized tests such as the National Assessment of Educational Progress (NAEP) and the National Educational Longitudinal Survey (NELS), other educational measures used for high-stakes decisions, such as high school grades, class rank, and indices of school quality and rigor of courses completed, as well as educational outcomes (e.g., college grades, persistence, graduation) (Camara and Schmidt, under review). Disparities in test results reflect similar differences in other measures and criteria (e.g., job performance, college achievement, and grades) and may be indicative of earlier differences in opportunity to learn 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 more subjective assessments that are both used daily to make high-stakes decisions about individual students, and have been shown to have similar levels of disparate impact against protected groups. Specifically, the Test Standards state, "tests include standardized ability (aptitude and achievement) instruments, diagnostic and evaluative devices, interest inventories, personality inventories, and projective instruments a more appropriate choice among assessment devices and subsequent use will be facilitated if there is a reasonable comparability in the kinds of information available to users." In the Standards, three broad categories of test instruments **are covered** [emphasis added]: constructed performance tasks, questionnaires, and to a lesser extent, structured behavioral samples (pages. 3-4)." Related to this comment, we suggest that the Resource Guide be renamed to put added emphasis on Measures Used in Making High-Stakes Decisions (parallel to the *Uniform Guidelines and Employment Selection Procedures*), rather than focus exclusively on just one element of the decision-making process, testing.

Fourth, we applaud OCR's deference to the Test Standards. However, the Resource Guide implies that existing professional standards can be applied in a rigid manner in evaluating tests. The Test Standards caution against such a rigid checklist approach, noting that specific circumstances affect the relevance of standards and professional judgment must be applied in evaluating tests. Professional practice and standards are typically construed more broadly and tests or other measures need not meet all standards to be appropriately used within the bounds of professional standards (Richardson, 729 F. Supp. At 821, 823). In addition, the three sponsoring educational associations are currently revising the Standards, which date back to 1985. We strongly endorse the recommendations from APA, AERA, and NCME to OCR asking that issuance of this Resource Guide be deferred until after publication and dissemination of the revised Test Standards and requesting a standard 90-day review period for any subsequent drafts of this document following publication of the revised Test Standards.

Fifth, we would ask OCR to ensure that colleges and universities, school districts, and state education agencies are given an opportunity to review and comment on this proposed Resource Guide. The Resource Guide has not been disseminated or reviewed by colleges and secondary schools. These are the very organizations that will be most directly affected by the Resource Guide

once it is issued and it seems appropriate that they be given an opportunity to review and comment on the inferences and proposed standards.

Sixth, the distinction the Resource Guide makes between tests and other assessment devices creates a false dichotomy, establishing a much lower technical, professional, and legal standard for more subjective assessment devices (e.g., applications, grades and GPA, recommendations, ratings or evaluations of student work and accomplishments, previous 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 (APA, 1987) that argued there is no professional or scientific justification to treat subjective and objective devices differently in imposing validation requirements. In fact, not imposing essentially the same legal and technical standards for all types of measures and devices used in high-stakes individual decisions would provide a sanctioned and covert means for discrimination. APA further argued that subjective procedures (in that case used for employment) are "amenable to the same psychometric scrutiny" as objective procedures, citing the Test Standards which address interviews and rating scales (Camara, 1996). In deciding *Watson v. Fort Worth Bank and Trust*, 487 U.S. 977, all eight of the justices joined Justice O'Connor's opinion holding that the adverse impact theory can be used in cases involving subjective practices. The Court was concerned that an employer could combine an objective criteria (such as a test or diploma) with subjective practices (such as interviews or ratings) and easily insulate itself from the Griggs test. O'Connor noted that "undisciplined" decision-making systems could have "precisely the same effects as a system pervaded by impermissible intentional discrimination (Opinion at 4926).

Seventh, professional and legal standards do not provide any support for OCR's distinctions between standardized tests and other measures. We agree with comments to an earlier draft of this Resource Guide submitted by the Board of Testing and Assessment (Shavelson, June 10, 1996), stating that "OCR's inquiry is not to pronounce judgment on the validity of inferences and decisions based on tests, but rather to determine whether the entire process of which the test is a part provides students a fair and equal opportunity to learn..." The Resource Guide ignores all factors other than tests even if they contribute more to disparate outcomes. In fact, high school courses, judgments about the relative "quality of high school curriculum," grades, and rank may also contribute more to disparate outcomes, than results from standardized testing, if an institution places substantially greater weight on these factors. For example, if tests are appropriately used as one of several factors in admissions, then there is no guarantee those disparate outcomes will be reduced if they are eliminated. In requiring tests to meet an exceptionally higher standard than other measures (GPA, ratings, samples of student work, high school rank, past experience, and opportunities), the Resource Guide will reduce or eliminate the use of valid and objective standardized tests used by educational institutions, states, and school districts. It is quite likely that educational institutions may opt to employ less valid and less objective methods for high-stakes decisions because they are not addressed in this Resource Guide.

Eighth, the Resource Guide also sanctions the use of the *Uniform Guidelines on Employment Selection Procedures* as a resource in educational testing. As the Resource Guide acknowledges in a footnote, there are critical and contextual differences between educational and employment testing that we believe undermines any attempt to use these guidelines in educational settings. The *Uniform Guidelines* were never developed with application to educational testing in mind and organizations did not have an opportunity to comment on extensions of the principles to education. In addition, the *Guidelines* are over 25 years old and do not reflect current scientific principles of measurement or current professional practice. The *Uniform Guidelines* are outdated and do not conform to the Testing Standards (AERA, APA, NCME, 1985) in their consideration of validity (as accomplished by adopting one of three distinct types of validity), validity generalization (this is virtually ignored in the *Guidelines*, but is accepted professional practice), differential prediction and classification, as well as several other areas (APA, 1985). The *Uniform Guidelines* may provide a framework for the development of guidelines addressing test use, but they should not be viewed as a substantive resource in educational assessment.

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

Tenth, this Resource Guide implies that once disparate impact is established that the burden is maintained by the educational institution to demonstrate both the educational necessity of the test and then to demonstrate that no alternative exists throughout the process. This legal interpretation is incorrect.

Other sections of the Resource Guide viewed as problematic include wording implying that separate local validation 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 setting cut scores 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 their current plan is to recirculate the current draft to groups who have already submitted comments on the current document. They will then submit a revised Guide to the National Academy of Sciences Board of Testing and Assessment, for its final review. Thereafter, they anticipate making a final draft available for public review this fall. They will publish a federal register notice 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 Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

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

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

Camara, W. (1996). Fairness and public policy in employment testing: Influences from a professional association. In, R. S. 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 testing, and other 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).

October 1999 [Table of Contents](#) | [TIP Home](#) | [SIOP Home](#)