

FOCAL ARTICLE

The *Uniform Guidelines* Are a Detriment to the Field of Personnel Selection

MICHAEL A. MCDANIEL, SVEN KEPES, AND GEORGE C. BANKS

Virginia Commonwealth University

Abstract

The primary federal regulation concerning employment testing has not been revised in over 3 decades. The regulation is substantially inconsistent with scientific knowledge and professional guidelines and practice. We summarize these inconsistencies and outline the problems faced by U.S. employers in complying with the regulations. We describe challenges associated with changing federal regulations and invite commentary as to how such changes can be implemented. We conclude that professional organizations, such as the Society for Industrial and Organizational Psychology (SIOP), should be much more active in promoting science-based federal regulation of employment practices.

For most of the history of the United States, the employment opportunities of ethnic and racial minorities, women, and older adults were substantially restricted. With the enactment of federal civil rights legislation, the U.S. government sought to end such employment discrimination. The *Uniform Guidelines on Employee Selection Procedures* (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice, 1978), hereafter "*Uniform Guidelines*," are U.S. federal guidelines, "which are designed to assist employers [...] to comply with requirements of federal law prohibiting employment practices which

discriminate on grounds of race, color, religion, sex, and national origin. They are designed to provide a framework for determining the proper use of tests and other selection procedures" (Section 1B). The *Uniform Guidelines* evolved from federal legislative actions and court decisions related to employment discrimination in the United States. As such, these 33-year-old guidelines have substantial influence on how employers, industrial and organizational (I–O) psychologists, and other practitioners in personnel selection conduct their work.

In this article, we present arguments that the *Uniform Guidelines* are scientifically inaccurate and inconsistent with professional practice as summarized in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), hereafter "*Standards*," and the *Principles for the Validation and Use of Personnel Selection Procedures*

Correspondence concerning this article should be addressed to Michael A. McDaniel.
E-mail: mamcdani@vcu.edu

Address: Virginia Commonwealth University, Snead Hall, 301 W. Main St., PO Box 844000, Richmond, VA 23284-4000

This paper has benefited substantially from the feedback of several individuals. Their help has been appreciated.

(SIOP, 2003), hereafter "*Principles*." We use these arguments to conclude that the *Uniform Guidelines* should be rescinded or at least extensively revised to be made consistent with current scientific knowledge and professional practice.

Encouraging Debate for the Betterment of Personnel Selection Practice

A discussion of the *Uniform Guidelines* is, in part, a discussion of mean racial differences. Past high-profile examinations of race-related issues (e.g., Herrnstein & Murray, 1994; Jensen, 1969) have been highly emotive. Within I–O psychology, the discussion of race is embedded in papers addressing high-stakes testing as well as personnel selection and job performance (e.g., McKay & McDaniel, 2006; Roth, Bevier, Bobko, Switzer, & Tyler, 2001; Sackett, Schmitt, Ellingson, & Kabin, 2001; Schmitt & Quinn, 2010), and these topics can also arouse emotion. In our experience, these topics tend not to be discussed in an open and professional manner and may degenerate into *argumentum ad hominem*, such as asserting that researchers who study demographic mean differences or who are critics of the *Uniform Guidelines* are racists, sexists, ageists, or are unsupportive of equal employment opportunity.

We note that nothing in our arguments for rescindment or extensive revision of the *Uniform Guidelines* is contrary to the authors' full support of equal employment opportunity. Nor are the arguments contrary to affirmative action or diversity efforts. Furthermore, the authors are strong advocates of continued research in understanding and reducing demographic mean differences in personnel selections tests and in assessments of job performance.

By presenting our arguments for the rescindment or revision of the *Uniform Guidelines*, we are hoping to foster a professional and collegial debate. Our article draws in part from previous work that either critiques the *Uniform Guidelines* or highlights differences between the *Uniform*

Guidelines and the *Standards* and/or *Principles* (e.g., Biddle, 2010; Cascio & Aguinis, 2001; Copus, 2006; Daniel, 2001; Ewos & Guseh, 2001; Jeanneret, 2005; Kleiman & Faley, 1985; McDaniel, 2007, 2010; O'Boyle & McDaniel, 2008; Sharf, 2006, 2008). We suggest that the lack of professional debate concerning the *Uniform Guidelines* damages the profession of I–O psychology by encouraging the use of personnel selection practices unsupported by scientific evidence. The lack of debate also encourages the gerrymandering of personnel selection practices (McDaniel, 2009), and a general disregard of the ethics of such practices. Furthermore, we suggest that the continued inaction of our professional organizations (e.g., SIOP) with respect to the inconsistency of the *Uniform Guidelines* with scientific knowledge and professional practice is unwise.

We begin the article with the assertion that the authoring agencies of the *Uniform Guidelines* made unfulfilled promises to keep the *Uniform Guidelines* and their interpretation consistent with scientific knowledge and professional practice. We then review sections of the *Uniform Guidelines* that are most disparate with scientific knowledge and professional practice. We offer evidence concerning the prevalence of racial disparities in employment screening results and suggest that these disparities should not generally trigger federal interference in personnel selection practices. We offer examples of how science and federal regulatory agencies interact. Finally, we call on the authoring agencies of the *Uniform Guidelines* to initiate a revision and provide suggestions for how SIOP and other professional organizations can encourage science-based federal regulation of employment practices.

The Unfulfilled Promises of the *Uniform Guidelines*

There is precedent for the revision of federal regulations related to employee selection. Before the *Uniform Guidelines* were

issued, the EEOC released employment testing regulations in 1966 (*Guidelines on Employment Testing Procedures*) and in 1970 (*Guidelines on Employee Selection Procedures*). The U.S. Civil Service Commission, the Department of Labor, and the Department of Justice had guidelines for similar purposes (Daniel, 2001). The issuance of successive guidelines may be viewed as an effort to maintain consistency with federal court decisions and scientific knowledge (Daniel, 2001). To avoid confusion among the differing guidelines issued by the four governmental agencies, the *Uniform Guidelines* were jointly issued in 1978 by the four agencies. They asserted that the *Uniform Guidelines* were intended to be consistent with professional practice and scientific findings. Specifically, in a section titled "Guidelines are consistent with professional standards," the *Uniform Guidelines* state:

The provisions of these guidelines relating to validation of selection procedures are intended to be consistent with generally accepted professional standards for evaluating standardized tests and other selection procedures, such as those described in the Standards for Educational and Psychological Tests prepared by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education (American Psychological Association, Washington, D.C., 1974) (hereinafter "A.P.A. Standards") and standard textbooks and journals in the field of personnel selection. (Section 5C)

The *Uniform Guidelines* also asserted that new scientific findings would be evaluated. In Section 5A, they state that "new strategies for showing the validity of selection procedures will be evaluated as they become accepted by the psychological profession." The *Uniform Guidelines*, when published in the *Federal Register*, included Supplementary Information, which includes the statement: "Validation has become

highly technical and complex, yet is constantly changing [...] Once the guidelines are issued, they will have to be interpreted in light of changing factual, legal, and professional circumstances" (p. 38292). With respect to construct validity, it is stated that the "guidelines leave open the possibility that different evidence of construct validity may be accepted in the future, as new methodologies develop and become incorporated in professional standards and other professional literature" (p. 38295). Thus, the agency authors of the *Uniform Guidelines* indicated that the guidelines and their interpretation should recognize advances in scientific knowledge and professional practice.

Scientific Knowledge, Professional Practice, and the Uniform Guidelines

Unfortunately for those who work in personnel selection and for the U.S. employers to whom they provide services, the authoring agencies of the *Uniform Guidelines* have failed to keep their promises to maintain and update the *Uniform Guidelines*. Thus, the next sections examine aspects of the *Uniform Guidelines* that substantially deviate from scientific knowledge and professional practice, ranging from the *Guidelines'* view of the situational specificity hypothesis to the lack of acknowledgement of the diversity–validity dilemma.

The Uniform Guidelines Embrace the Situational Specificity Hypothesis

Beginning in the 1920s and continuing into the 1970s, it was observed that the same employment test yielded different validity results across settings (Schmidt & Hunter, 1998). For example, a test to screen bank tellers in one bank would yield a high validity (i.e., a high magnitude correlation between the test and job performance) but could yield a much lower validity for bank tellers in a bank across the street. Such findings were frequent and led to speculation that there were as yet undiscovered

characteristics of employment situations that caused a test to be valid for one location but not for another. This speculation became known as the situational specificity hypothesis, which was widely accepted as fact (Guion, 1975; Schmidt & Hunter, 2003).

Given that the situational specificity hypothesis suggested that there were unknown causes of validity differences despite apparently similar employment situations and jobs, professional practice emphasized the conduct of detailed job analyses. There was an assumption that conducting detailed job analyses would uncover differences among employment situations that caused validities to vary across similar situations and jobs. As knowledge of the validity of a test in one situation for a given job did not always predict the validity of the same test in a similar situation and job, professional practice emphasized conducting local validation studies. Consistent with this thinking, the *Uniform Guidelines* emphasized the practices of detailed job analyses and local validation studies.

Beginning in 1977, Schmidt and Hunter began publishing empirical evidence discrediting the situational specificity hypothesis. Specifically, they demonstrated that much of the variability in validity coefficients across studies was due to random sampling error. Any primary study examining the correlation between a test and job performance seeks to estimate the validity coefficient in the population. When sample sizes are relatively small (e.g., $N < 500$), the samples have a high probability of being nonrepresentative of the population and thus likely to offer an imprecise estimate of the population validity. Thus, the validity coefficient derived from a small sample might over- or underestimate the population validity. At the time of Guion's classic text (Guion, 1965), the average sample size in a validity study was 68. We now know that this sample size is far too small to estimate the true validity of a test in the population accurately. For instance, a test with a population validity of .20 could easily yield sample validities ranging from $-.04$

to $.42^1$ based on sample sizes of 68. Thus, small sample studies make validity coefficients appear unstable even when they are constant in the population.

The Emphasis of the Uniform Guidelines on Local Validation Studies

The *Uniform Guidelines* require validity evidence when a test demonstrates adverse impact (i.e., differential hiring rates by race, gender, etc.). Yet, for most employers, local empirical validity studies are professionally ill-advised because of sample-size limitations. In contrast, the *Uniform Guidelines* are largely oblivious to sample size issues in test validation. The *Principles* acknowledge that "validation planning must consider the feasibility of the design requirements necessary to support an inference of validity. Validation efforts may be limited by time, resource availability, sample size, or other organization constraints including cost" (p. 10). From the perspective of precision in estimating a population validity coefficient, sample sizes below 100 are clearly inadequate, yet 79% of U.S. employers have fewer than 100 employees and 84% have fewer than 500 employees (U.S. Census Bureau, 2007). The employees of these small- to medium-sized businesses would likely be found in multiple occupations, further reducing the sample size available for a concurrent validation study of a single occupation. Similarly, such small employers are likely to hire relatively few employees in a given time period, making predictive validity studies infeasible as well. In brief, only a small percentage of employers have enough employees in a given occupation to permit credible local criterion-related validity documentation. Thus, with respect to criterion-related validity evidence, the *Uniform Guidelines* seek documentation that

1. A point estimate of .2 with a sample size of 68 leads to a 95% confidence interval ranging from $-.04$ to $.42$.

cannot be provided by the majority of U.S. employers.²

The Uniform Guidelines and Evidence for Validity Based on Content Similarity

We note that both the *Principles* and the *Uniform Guidelines* address standards for validity documentation.³ However, the *Uniform Guidelines* adopted a curious stance with respect to what job-related personal characteristics can and cannot be defended based on content evidence. Without any stated science-based justification, the *Uniform Guidelines* declare:

A selection procedure based upon inferences about mental processes cannot be supported solely or primarily on the basis of content validity. Thus, a content strategy is not appropriate for demonstrating the validity of selection procedures which purport to measure traits or constructs, such as intelligence, aptitude, personality, commonsense, judgment, leadership, and spatial ability. (Section C1)

We note that this section of the *Uniform Guidelines* appears to rule out a content validity defense for some very common selection constructs including general and specific tests of cognitive ability and the Big Five personality traits. It would also appear to exclude content validity as a defense for most interviews, assessment centers, and situational judgment tests to the extent that the measures seek to assess constructs associated with cognitive ability, personality, and leadership.⁴ This situation leaves most U.S.

employers in a very bad situation because few employers have sufficient employees or applicants to conduct a criterion-related validity study, and they are further precluded from using a content validity strategy to defend reasonable tests of cognitive ability or personality.

The *Uniform Guidelines* do not appear to appreciate problems created in organizations as a result of the regulation. For example, the *Uniform Guidelines* approach to content validity is problematic for many organizations with rapidly evolving work and flexible occupational structures. In contrast, the *Principles* note that organizations experiencing “rapid changes in the external environment, the nature of work, or processes for accomplishing work may find that traditional jobs no longer exist. In such cases, considering the competencies or broad requirements for a wider range or type of work activity may be more appropriate” (p. 9). In addition, the *Principles* note the value of a less detailed approach to job analysis than is found in the *Uniform Guidelines*:

A less detailed analysis may be appropriate when prior research about the job requirements allows the generation of sound hypotheses concerning the predictors or criteria across job families or organizations. When a detailed analysis of work is not required, the researcher should compile reasonable evidence establishing that the job(s) in question are similar in terms of work behavior and/or required knowledge, skills, abilities, and/or other characteristics, or falls into a group of jobs for which validity can be generalized. (p. 11)

-
2. We note that this requirement from the *Uniform Guidelines* has led to consortium groups (e.g., Edison Electric Institute and Mayflower) that conduct industry-wide selection validation studies. However, although these consortiums are useful to a few large industries (e.g., electric utilities), they have limited applicability to many U.S. employers.
 3. We have some concerns regarding the use of the *Uniform Guidelines* as a cookbook for job analysis. However, these concerns are criticisms of job analysts and not so much the *Uniform Guidelines*.
 4. We recognize that content validity documentation in practice is often offered for mental constructs

and measurement methods such as assessment centers. This is done in part by changing what one calls constructs. Thus, an employment test assessing intelligence (i.e., general cognitive ability) by a composite of three ability tests (reading comprehension, numerical fluency through tables, and reasoning) would be presented as the following attributes: ability to read, ability to work with tables, and ability to solve problems.

We assert that cost and time constraints make the *Uniform Guidelines'* content validity requirements burdensome for many U.S. employers because they may lack the financial and technical resources to fully comply with them. The *Principles* address feasibility limitations on job analysis for content validity: "Among these issues are the stability of the work and the worker requirements, the interference of irrelevant content, the availability of qualified and unbiased subject matter experts, and cost and time constraints" (p. 21).

The Uniform Guidelines and Evidence for Validity Based on Construct Validity

The *Standards* state that validation begins with "an explicit statement of the proposed interpretation of test scores, along with a rationale for the relevance of the interpretation to the proposed use. The proposed interpretation refers to the constructs or concepts the test is intended to measure" (p. 9). Thus, although all validation concerns constructs, the *Uniform Guidelines* adopted a curious position concerning construct approaches to validity evidence:

Construct validity is a more complex strategy than either criterion-related or content validity. Construct validation is a relatively new and developing procedure in the employment field, and there is at present a lack of substantial literature extending the concept to employment practices. The user should be aware that the effort to obtain sufficient empirical support for construct validity is both an extensive and arduous effort involving a series of research studies, which include criterion related validity studies and which may include content validity studies. Users choosing to justify use of a selection procedure by this strategy should therefore take particular care to assure that the validity study meets the standards set forth below. (Section D1)

This wording made it largely impossible to use construct evidence as a validity

defense under the *Uniform Guidelines*. Counter to the statement in the Supplementary Information (p. 38295) of the *Uniform Guidelines* concerning the evaluation of new scientific approaches to construct validity, the *Uniform Guidelines* have never been revised with respect to construct validity.

In contrast to the nonscientific assertions of the *Uniform Guidelines*, the *Principles* and *Standards* recognize the importance of varied approaches to construct evidence in support of validity. The *Principles* highlight the value of validity evidence demonstrating the relationship between an employment test and other variables. For example, the *Principles* state that "evidence that two measures are highly related and consistent with the underlying construct can provide convergent evidence in support of the proposed interpretation of test scores as representing a candidate's standing on the construct of interest" (p. 5). The *Principles* also discuss the usefulness of discriminant validity and the value of evidence relating to the internal structure of the test. For example, a high degree of item internal consistency would be supportive of a test argued to represent a single construct.

The Uniform Guidelines and Its 1950s Perspective on Separate "Types" of Validity

The *Principles* note that in the early 1950s three different types of test validity were considered, these being content, criterion-related, and construct. The measurement literature has since adopted the perspective that validity is a unitary concept in which different sources of information can inform inferences about test scores. The *Principles* emphasize that "nearly all information about a selection procedure, and inferences about the resulting scores, contributes to an understanding of its validity. Evidence concerning content relevance, criterion relatedness, and construct meaning is subsumed within this definition of validity" (p. 4). In contrast to the professional

practice summarized in the current *Principles and Standards*, the *Uniform Guidelines* continue to embrace the 1950s perspective on three distinct types of validity.

The Uniform Guidelines and Meta-Analysis As a Source of Validity Documentation

The early work of Schmidt, Hunter and colleagues (e.g., Pearlman, Schmidt, & Hunter, 1980; Schmidt, Gast-Rosenberg, & Hunter, 1980; Schmidt & Hunter, 1977) concerning situational specificity evolved into psychometric meta-analysis procedures (Hunter & Schmidt, 2004). The application of meta-analysis to validity data became known as validity generalization, and a test was argued to show validity generalization when a large majority (typically 90%) of population validities were above 0. The *Standards* and the *Principles* endorse validity generalization as evidence of the validity of employment tests. The *Principles*, for instance, note:

Meta-analysis is the basis for the technique that is often referred to as "validity generalization." In general, research has shown much of the variation in observed differences in obtained validity coefficients in different situations can be attributed to sampling error and other statistical artifacts (Ackerman & Humphreys, 1990; Barrick & Mount, 1991; Callender & Osburn, 1980, 1981; Hartigan & Wigdor, 1989; Hunter & Hunter, 1984; Schmidt, Hunter, & Pearlman, 1981). These findings are particularly well established for cognitive ability tests; additional recent research results also are accruing that indicate the generalizability of predictor-criterion relationships for noncognitive constructs in employment settings. (p. 28)

From the perspective of scientific knowledge, meta-analytic evidence largely eliminates the need for local validity studies. Specifically, only if "important conditions in the operational setting are not represented in the meta-analysis (e.g., the local

setting involves a managerial job and the meta-analytic data base is limited to entry level jobs)" do the *Principles* state that local individual studies "may be more accurate than the average predictor-criterion relationship reported in a meta-analytic study" (p. 29). In addition to the acceptance of validity generalization in professional standards, courts have found in favor of generalizing validity evidence (see Sharf, 2006).

We recognize that most of the evidence concerning validity generalization was developed after the publication of the *Uniform Guidelines*. However, the *Uniform Guidelines* have never been revised to acknowledge the role of meta-analysis in demonstrating the validity of employment tests. Reliance on validity generalization evidence may be one of the most economical approaches to test validation, and its omission from the *Uniform Guidelines* is inappropriate.

We speculate that a primary reason why the *Uniform Guidelines* have not been revised to incorporate validity generalization as an acceptable validity defense is that it might change the litigation landscape significantly. There are concerns that assessments with strong validity generalization support, such as general cognitive ability, will become more widely used and result in a less racially diverse workforce. There are also individuals and organizations, such as employment attorneys, expert witnesses, employment testing consultants, and enforcement agencies, whose business is driven, in part, by the *Uniform Guidelines*. If litigation becomes less frequent due to a wider acceptance of validity generalization as a validity defense, some individuals and organizations will suffer financial harm. Finally, there are some who are worried that validity generalization could be applied inappropriately as a validation defense. This concern could be reduced by more guidance, such as is found in the *Principles*, concerning how validity generalization results may be applied appropriately to specific testing situations (Banks & McDaniel, in press; McDaniel, 2007).

The Uniform Guidelines and Restrictions on Transportability of Evidence

Although applications of meta-analysis to validity data may be viewed as transportability of evidence supporting validity, the use of the word transportability often refers to using information from a primary validity study to generalize validity to the use of the test in a new situation. The *Principles* address the value of transportability evidence in the documentation of the validity of employment tests:

One approach to generalizing the validity of inferences from scores on a selection procedure involves the use of a specific selection procedure in a new situation based on results of a validation research study conducted elsewhere. This is referred to as demonstrating the “transportability” of validity evidence for the selection procedure. When proposing to “transport” use of a procedure, a careful review of the original validation study is warranted to ensure acceptability of the technical soundness of that study and to determine its relevance to the new situation. Key points for consideration when establishing the appropriateness of transportability are, most prominently, job comparability in terms of content or requirements, as well as, possibly, similarity of job context and candidate group. (p. 26)

We note that the transportability language in the *Principles* does not limit the type of validity evidence. Unfortunately, in the *Uniform Guidelines*, transportability is only mentioned with respect to criterion-related validity. With respect to content validity, a reviewer has advised us that the “transport” of content evidence devolves to the job analysis and demonstration of the job relevance of the content, effectively repeating the content evidence from the original study. In brief, the *Uniform Guidelines* make transportability of validity evidence based on content or construct relevance a difficult proposition and thus

are, once again, inconsistent with scientific knowledge and professional guidelines.

The Uniform Guidelines Position With Respect to Differential Validity and Differential Prediction

Belief in the situational specificity hypothesis coupled with the very common observation of mean racial differences in test scores encouraged scientific inquiries regarding the possibility of differential validity and differential prediction (Boehm, 1977; Bray & Moses, 1972; Kirkpatrick, Ewen, Barrett, & Katzell, 1968). It was argued that the validity (i.e., differential validity) or the prediction accuracy (i.e., differential prediction) may vary by ethnic and racial group. However, during the late 1970s and early 1980s, it became evident that differential validity was rare (Schmidt, 1988; Schmidt & Hunter, 1981; Wigdor & Garner, 1982). Differential prediction might result from either differing slopes or differing intercepts. By the late 1970s, it was demonstrated that differential prediction by slope does not occur at higher levels than expected by chance (Bartlett, Bobko, Mosier, & Hannan, 1978). Differential prediction by intercept is less rare, but the error in prediction tends to favor minority groups (Hartigan & Wigdor, 1989; Schmidt, Pearlman, & Hunter, 1980).

Unfortunately, the most definitive scientific knowledge concerning differential validity and prediction developed largely after the publication of the *Uniform Guidelines*. However, already in 1978, many I–O psychologists believed that differential prediction did not exist (Daniel, 2001; Hunter, Schmidt, & Hunter, 1979). Thus, the differential prediction requirement in the *Uniform Guidelines* may have been included due to enforcement considerations rather than technical or scientific knowledge (Daniel, 2001). Nevertheless, even with the accumulation of scientific knowledge concluding that “differential validity does not exist” (Gatewood, Feild, & Barrick, 2008, p. 547) and that differential prediction typically does not occur, and when it

does, it tends to favor minority groups (Hartigan & Wigdor, 1989; Schmidt, Pearlman, & Hunter 1980), the *Uniform Guidelines* have not been revised to be consistent with current knowledge.

We note the recent resurgence of scientific interest in differential prediction (Aguinis, Culpepper, & Pierce, 2010; Borneman, 2010; Meade & Tonidandel, 2010). As with all areas concerning personnel selection and equal employment opportunity, we encourage continued research. For our discussion, we suggest that the most relevant aspect of this research concerns statistical power. Given that research generally argues that differential prediction studies are almost always underpowered, it makes little sense for the *Uniform Guidelines* to encourage differential prediction studies when the sample sizes available to the vast majority of employers are too small to detect whether differential prediction exists. This is yet one more area where the *Uniform Guidelines* are inconsistent with current scientific knowledge.

The Uniform Guidelines and False Assumptions Concerning Adverse Impact

The *Uniform Guidelines* incorporate the 4/5ths rule to determine if adverse impact is present. If the ratio of the minority hiring rate is less than 80% of the majority hiring rate, adverse impact is generally considered present. We note that the 4/5ths rule has no scientific basis, and there are debates concerning its value (Cohen, Aamodt, & Dunleavy, 2010; Roth, Bobko, & Switzer, 2006; Shoben, 1978). Although not mentioned in the *Uniform Guidelines*, federal enforcement agencies often use a "two standard deviation test," which is a statistical test for differences in proportions. Both the 4/5ths rule and the "two standard deviation test" have been criticized as techniques for assessing adverse impact (Morris & Lobsenz, 2000; Roth et al., 2006). When hiring decisions result in adverse impact, the *Uniform Guidelines* make it the responsibility of the employer to provide test validation documentation. Developing such

documentation can be very expensive and labor intensive because it often requires the service of consulting firms, expert witnesses, and other specialists. Although we are not arguing that such validation evidence is not desirable for all selection procedures, compliance with the *Uniform Guidelines* documentation requirements can prove to be very expensive, particularly for small- and medium-sized employers who comprise the large majority of U.S. employers.

We suggest that an implicit assumption of the *Uniform Guidelines* is that adverse impact is an indication of a flawed test. We offer the alternative hypothesis that the employment test is an accurate assessment of subgroup differences in job-related attributes. Table 1 summarizes the field's cumulative knowledge on the extent of mean score differences by race and gender. It is clear that almost all selection procedures, possibly excepting personality, are likely to show mean racial differences of sufficient magnitude to typically result in adverse impact for any reasonable passing point. Thus, unfortunately, adverse impact is the norm and not the exception. We argue that the common finding of mean racial differences and the potential causes of the mean racial differences in employment tests are "the elephant in the room" of personnel selection (i.e., a large and obvious problem that is seldom discussed). We also argue that, given the pervasiveness of adverse impact, the presence of adverse impact should not result in federal interference in employment practices when such interference is based on regulations inconsistent with scientific knowledge. Note that we are strong advocates that all selection procedures should be job related. What we object to is a requirement that validation evidence must comply with scientifically inappropriate federal regulations.

We offer that the primary cause of mean racial differences in employment test scores is mean racial differences in job-related attributes, not flawed employment tests. We suggest that employment tests are measuring mean racial differences in job-related attributes accurately. We offer the following

Table 1. Meta-Analytic Standardized Racioethnic and Gender Subgroup Differences and Validities

Predictor ^a	<i>d</i> -value(s)	Criterion-related validity
General cognitive ability		.51 ^b
White–Black	.99 ^b	
White–Hispanic	.58 to .83 ^b	
White–Asian	–.20 ^b	
Male–female	.00 ^b	
Conscientiousness		.18 ^b
White–Black	.06 ^b and .07 ^c	
White–Hispanic	.04 ^b and .08 ^c	
White–Asian	.08 ^b and .11 ^c	
Male–female	–.08 ^b	
Conscientiousness, global measures		
White–Black	.17 ^c	
White–Hispanic	.20 ^c	
White–Asian	.04 ^c	
Conscientiousness, achievement		
White–Black	–.03 ^c	
White–Hispanic	.10 ^c	
White–Asian	.14 ^c	
Conscientiousness, dependability		
White–Black	–.05 ^c	
White–Hispanic	.00 ^c	
White–Asian	–.01 ^c	
Conscientiousness, cautiousness		
White–Black	.16 ^c	
Conscientiousness, order		
White–Black	.01 ^c	
White–Hispanic	.00 ^c	
White–Asian	.50 ^c	
Extraversion		.11 ^b
White–Black	.10 ^b and –.16 ^c	
White–Hispanic	–.01 ^b and –.02 ^c	
White–Asian	.15 ^b and –.14 ^c	
Male–female	.09 ^b	
Extraversion, global measures		
White–Black	–.21 ^c	
White–Hispanic	.12 ^c	
White–Asian	–.07 ^c	
Extraversion, dominance		
White–Black	–.03 ^c	
White–Hispanic	–.04 ^c	
White–Asian	–.19 ^c	
Extraversion, sociability		
White–Black	–.39 ^c	
White–Hispanic	–.16 ^c	
White–Asian	–.09 ^c	
Emotional Stability		.13 ^b
White–Black	–.04 ^b and –.09 ^c	
White–Hispanic	–.01 ^b and .03 ^c	

Table 1. *Continued*

Predictor ^a	<i>d</i> -value(s)	Criterion-related validity
White-Asian	.08 ^b and $-.12^c$	
Male-female	.24 ^b	
Emotional Stability, global measures		
White-Black	$-.12^c$	
White-Hispanic	$-.04^c$	
White-Asian	$-.16^c$	
Emotional Stability, self-esteem		
White-Black	.17 ^c	
White-Hispanic	.25 ^c	
White-Asian	.30 ^c	
Emotional Stability, low anxiety		
White-Black	$-.23^c$	
White-Hispanic	.25 ^c	
White-Asian	.27 ^c	
Emotional Stability, even tempered		
White-Black	.06 ^c	
White-Hispanic	.09 ^c	
White-Asian	$-.38^c$	
Agreeableness		.08 ^b
White-Black	.02 ^b and $-.03^c$	
White-Hispanic	.06 ^b and $-.05^c$	
White-Asian	.01 ^b and .63 ^c	
Male-female	$-.39^b$	
Openness to Experience		.07 ^b
White-Black	.21 ^b and $-.10^c$	
White-Hispanic	.10 ^b and $-.02^c$	
White-Asian	.18 ^b and .11 ^c	
Male-female	.07 ^b	
Job knowledge		.48 ^b
White-Black	.48 ^b	
White-Hispanic	.47 ^b	
Spatial ability		.51 ^b
White-Black	.66 ^b	
Psychomotor ability		.35 ^b
White-Black	$-.72^d$	
White-Hispanic	$-.11^d$	
Male-female	-1.06^d	
Psychomotor ability, muscular strength		.23 ^b
Male-female	1.66 ^b	
Psychomotor ability, muscular power		.26 ^b
Male-female	2.10 ^b	
Psychomotor ability, muscular endurance		.23 ^b
Male-female	1.02 ^b	
Biodata		.35 ^b
White-Black	.33 ^b	
Structured interview		.51 ^b
White-Black	.23 ^b	

Table 1. *Continued*

Predictor ^a	<i>d</i> -value(s)	Criterion-related validity
Situational judgment test (SJT)		
Video SJT		.22 to .33 ^d
White–Black	.31 ^b	
White–Hispanic	.41 ^b	
White–Asian	.49 ^b	
Male–female	–.06 ^b	
Written SJT		.34 ^b
White–Black	.40 ^b	
White–Hispanic	.37 ^b	
White–Asian	.47 ^b	
Male–female	–.12 ^b	
Accomplishment record		.17 to .25 ^d
White–Minority	.24 ^d	
Male–female	.09 ^d	
Work sample		.33 ^b
White–Black	.52 ^b	
White–Hispanic	.45 ^b	
Assessment center		.37 ^b
White–Black	.60 or less ^d	

Note. Drawn from Ployhart and Holtz (2008) and Foldes, Duehr, and Ones (2008).

^aPredictors encompass predictor constructs that assess one construct (e.g., cognitive ability, Conscientiousness, and Extraversion) and predictor measurement methods that assess multiple constructs. For predictor measurement methods, the magnitude of group differences will be a function of the constructs assessed. For racial comparisons, a positive *d* indicates Whites score higher than the other group on average. For comparisons by gender, a positive *d* indicates men score higher than women on average. ^bEstimate from Ployhart and Holtz (2008); corrected unless otherwise indicated. ^cEstimate from Foldes, Duehr, and Ones (2008). ^dEstimate from Ployhart and Holtz (2008). Estimate is from primary studies; *not* meta-analytically derived.

lines of evidence in support of our position. First, mean differences are often substantial and present prior to the age in which people begin competing for jobs. For example, mean racial differences are found early in life (e.g., age 3; see Jencks & Phillips, 1998; Phillips, Brooks-Gunn, Duncan, Klebanov, & Crane, 1998). Clearly, mean racial differences at age 3 cannot be attributed to flawed employment tests.

In further support of our position, we describe two sources of data relevant to those currently in the workforce: high school graduation rate and prose literacy in U.S. adults. High school graduation rates by ethnicity and race are available from the National Center for Educational Statistics (Stillwell, 2010). In these data, high school graduation is defined as receiving a high school diploma at the completion of 4 years of high school for the cohort

graduating in the spring of 2008. Ninety-one percent of Asians, including Pacific Islanders, receive a high school diploma. Ten percent fewer (81%) of Whites receive one. For American Indians, including Alaskan natives, the diploma recipient rate is 64%, which is tied with the Hispanic rate. The percent of Blacks receiving a high school diploma is 62%. We assert that high school diploma status covaries with many job-related attributes, including general cognitive ability and Conscientiousness. Both of these attributes show validity generalizations for virtually all jobs (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Hunter, Schmidt, & Le, 2006; Hurtz & Donovan, 2000; Schmidt & Hunter, 1998).

In 2011, individuals in this cohort are approximately 22 years of age, and most of them are likely employed or competing for employment. These individuals are

also likely to be employed or apply for employment for the next 43 years, at which time they will reach the age of 65. We suggest that the job-related attributes associated with high school diploma status will likely yield adverse impact for this age cohort for the next 43 years. Former Supreme Court Justice O'Connor, in her majority opinion in the *Grutter v. Bollinger* (2003) case concerning racial preferences in law school admission, wrote: "We expect that 25 years from now, the use of racial preferences will no longer be necessary to further the interest approved today." We respectfully suggest that her opinion was not based on a realistic appraisal of available data. We offer an opinion based on science: Mean racial differences in educationally relevant and job-related attributes will, unfortunately, not go away any time soon.

Our second data set concerns prose literacy for a representative sample of U.S. adults for the year 2003 (National Center for Education Statistics, 2010). This data source defines an intermediate level of literacy as "able to read and understand moderately dense, less commonplace prose text, as well as summarize, make simple inferences, determine cause and effect, and recognize author's purpose" (National Center for Education Statistics, 2010, footnote 1). We offer that most knowledge-worker occupations require incumbents to read and understand moderately dense prose, to make simple inferences, and to determine cause and effect. We suggest that one typically needs these skills to graduate from high school. The 2003 data from the National Center for Educational Statistics indicate that 51% of Whites fall in this intermediate level of skills, compared to 42% for Asians, 31% percent for Blacks, and 23% for Hispanics. We suggest that until a time when mean racial differences in prose literacy are eliminated, regrettably, most valid employment tests are likely to have adverse impact.

We encourage educational and other interventions that would eliminate or reduce these mean racial differences in job-related attributes. However, we are not

hopeful that these differences will be eliminated any time soon. Part of our pessimism is based on the intervention research summarized by Ceci and Papierno (2005). Even if there was an intervention that would dramatically improve job-related attributes, we should not assume that such an intervention would close the achievement gap between the less able and the more able. Rather, the intervention might increase the gap, partly because the more able have a higher capacity to benefit more from the intervention and partly because the more able will be more likely to participate in the intervention (Ceci & Papierno, 2005; Walberg & Tsai, 1983). Thus, even with dramatically impressive interventions, mean racial differences may persist (Ceci & Papierno, 2005). Given the prevalence of mean racial differences, employers are typically in need of a validation defense consistent with federal regulations. Thus, it is imperative that federal regulations permit all scientifically based approaches to validity evidence. Currently, they do not.

The Uniform Guidelines and the Diversity-Validity Dilemma

The *Uniform Guidelines* are silent about the diversity-validity dilemma (Ployhart & Holtz, 2008; Pyburn, Ployhart, & Kravitz, 2008) that organizations face and how organizations should deal with this dilemma. When faced with the adverse impact of an employment test, the *Uniform Guidelines* encourage employers to search for alternative tests with the same or higher validity but less adverse impact. Such searches are almost always futile. Current employment tests seldom maximize diversity and validity goals because the validity of employment tests tends to co-vary with mean racial differences such that the most valid tests have the largest mean racial differences (Pyburn et al., 2008).

Organizations can use two strategies to deal with this diversity-validity dilemma (Pyburn et al., 2008). First, they can sacrifice validity and use less valid selection tests that do not result in adverse impact

to achieve social, ethical, or business aims.⁵ Second, organizations can *sacrifice diversity* by ignoring the potential adverse impact of valid selection procedures to achieve different social, ethical, or business aims. Obviously, neither strategy is optimal because the first can sacrifice work quality and utility (Hunter & Hunter, 1984; Schmidt & Hunter, 1998), and the second can result in racial imbalance and discrimination lawsuits. Thus, both strategies ultimately impinge on important social, ethical, and economic objectives (Pyburn et al., 2008). Although the scientific community has debated this issue and provided recommendations of how to deal with the dilemma (e.g., Kravitz, 2008; Ployhart & Holtz, 2008; Pyburn et al., 2008), the legality of some of the proposed solutions is not clear. Unfortunately, the *Uniform Guidelines* do not address this vital issue. Thus, they implicitly deny any dilemma or tradeoff.

The Broader Political and Social Context and the *Uniform Guidelines*

In the previous sections, we reviewed the inconsistencies between scientific knowledge and the *Uniform Guidelines*. Next, we speculate about the forces influencing the inertia of the *Uniform Guidelines* and present ideas about how they could be revised to reflect current scientific knowledge and professional practice.

Resistance to Changing the Uniform Guidelines

Despite the overwhelming evidence that the *Uniform Guidelines* are not in compliance with important legal, technical,

and scientific developments (Daniel, 2001; McDaniel, 2007), they have remained unchanged for over 3 decades. Table 2 summarizes inconsistencies between the *Uniform Guidelines* and science-based professional practice.

To address some of these issues, several attempts have been made to revise the *Uniform Guidelines*. For instance, the General Accounting Office proposed a review of the *Guidelines* in 1982 (Daniel, 2001). However, all efforts, including an oversight hearing on the Civil Rights Division of the U.S. Department of Justice and several hearings before the Committee on Education and Labor, Subcommittee on Employment Opportunities, regarding the *Uniform Guidelines* in 1985, yielded no tenable outcome (Daniel, 2001). Later, efforts in 1998 were equally fruitless (Daniel, 2001). A partisan political climate may have prevented a science-based revision of the *Uniform Guidelines*. We suggest that the best hope for the revision of the *Uniform Guidelines* lies with the Obama administration. Given President Obama's mixed-racial heritage, an Obama-endorsed congressional effort to force a revision of the *Uniform Guidelines* is less likely to be labeled as racially motivated.

The Role of Science in Federal Regulations

The failure to maintain the *Uniform Guidelines* consistent with science and professional practice is unfortunate. Other federal laws and regulations are updated regularly to address new scientific evidence. For instance, consumer protection would have suffered if Congress had not passed the Food and Drug Administration Amendments Act of 2007. Similarly, businesses, potential applicants, current employees, and the I-O psychology profession are not well served by federal employment guidelines that are inconsistent with legal, technical, and scientific developments.

We believe that the appropriate role of science in federal employment regulations can be explored by examining nonemployment regulatory areas. Across

5. We acknowledge that a combination of a cognitive ability test and a noncognitive measure may improve the validity to some degree while reducing adverse impact to some extent. Our reading of the literature causes us to conclude that the improvements in validity and the reductions in adverse impact, when occurring, are typically relatively modest. Thus, the use of such composites provides, at best, only a limited reduction of the problems associated with the validity-diversity dilemma.

Table 2. Summary of Scientific and Practical Problems and Inconsistencies in the Uniform Guidelines

Problem/ inconsistency	Uniform Guidelines	Scientific knowledge and professional practice
General		
Issue date	1978	1999 (<i>Standards</i>) and 2003 (<i>Principles</i>)
Scientific/practical		
Situational specificity hypothesis	Endorsement of the situational specificity hypothesis	Rejection of the situational specificity hypothesis
Local validation studies	Requirement of local validation studies	No requirement of local validation studies
Content validity evidence	Rejection of content validity evidence-based defense strategies for some constructs	No construct limitations on what can be validated
Construct validity assessment	Practical rejection of construct validity evidence-based defense strategies	Practical endorsement of construct validity evidence-based defense strategies
View of validity	Outdated perspective of the concept of validity (i.e., there are three distinct types of validity)	Endorsement of validity is a unitary concept in which different sources of information can inform inferences about a selection approach
Validity generalization	Outdated perspective on validity generalization as evidence for the validity of employment tests	Endorsement of validity generalization as evidence of the validity of employment tests
Transportability of evidence	Transportability may only apply to criterion-related validity	Transportability applies to the concept of validity as a whole
Differential validity and differential prediction	Requirement of the assessment of differential validity and prediction evidence	Differential validity is unlikely to exist; no assessment is necessary
Assumptions concerning adverse impact	A flawed employment test leads to adverse impact	Multiple causes could lead to adverse impact
The diversity–validity dilemma	No clear guidance	Guidance is provided

scientific areas, from educational interventions to environmental protection and medical research, powerful economic and social interests are often at play (Steinbrook, 2004). Political entities can be driven to influence science for both economic and social reasons. However, scientific evidence is not an à la carte menu from which policymakers should be able to selectively pick popular research and avoid results which are unpopular (Schenkel, 2010). It is critical that a clear distinction be made

between honest scientifically based challenges and politically motivated attacks on scientific evidence (Rosenstock & Lee, 2002). To assist in this distinction, one must first recognize the influence tactics often used, including economic, manufacturing uncertainty, and delay tactics (for a good overview of the influence and impact of such tactics see Rosenstock & Lee, 2002). As a result of such tactics, federal regulations can be delayed and misguided, which can result in uncertainty, financial

and economic loss (Michaels & Monforton, 2005; Rosenstock & Lee, 2002; Slavin, 2002), as well as human loss as was the case when regulation requiring a simple warning label on aspirin bottles indicating that aspirin could increase children's risk of Reye's syndrome was successfully delayed by the aspirin industry (Michaels & Monforton, 2005).

We suggest that all three tactics (e.g., economic, manufacturing uncertainty, and delay) will be used both for and against efforts to make the *Uniform Guidelines* consistent with scientific evidence and professional practice. First, employers can document the costs associated with complying with the *Uniform Guidelines*. These include labor and other monetary costs associated with defending employee selection systems. There are also economic costs associated with using lower validity selection measures in hopes of reducing adverse impact (Hunter & Hunter, 1984; Schmidt & Hunter, 1998). Second, employees of federal regulatory agencies, human resources consultants, and labor lawyers seeking to preserve their jobs can manufacture uncertainty about scientific findings. If the price is right, one can find a "scientist" to testify to almost anything. Third, regulatory agencies and other interested parties (e.g., consultants, lawyers, and expert witnesses) can engage in delay tactics (e.g., litigation, requiring parallel studies and fighting over access to raw data) to avoid revising the *Uniform Guidelines*. Some might argue that delay tactics have contributed to the fact that no revisions have been made to the *Uniform Guidelines* in over 3 decades.

Changing Federal Regulations Concerning Employment Testing

The rescindment or revision of the *Uniform Guidelines* faces a variety of obstacles. First, employers may not like the *Uniform Guidelines* and the expense of complying with them, but they tend to like stability. Changes in the federal regulation of employment practices create uncertainty, which may not be welcome by many employers. Second,

courts have given deference to the *Uniform Guidelines* in hundreds of cases, and courts generally abide by precedent. Thus, courts may be unlikely to alter their practices to be consistent with scientific knowledge without changes to existing federal law such as the Civil Rights Act of 1991. In addition, even if the *Uniform Guidelines* were revised to be consistent with scientific knowledge, there would still be a need to influence and alter a formidable body of case law. Third, there are political obstacles to acknowledging that adverse impact could reflect mean racial differences in job-related attributes and that the mean racial gap in such attributes is not going away any time soon. It is easier for Congress, the courts, and regulatory agencies to encourage the belief that employment tests with adverse impacts are likely flawed than to admit that there are mean racial differences in job-related attributes. However, based on trends in the debates of educational testing, we have some hope that these organizations can accept conclusions based on clear data. In K-12 educational testing, there was once substantial debate concerning "biased tests." With the passing of the No Child Left Behind Act in 2001, there appears to be an implicit acceptance of the conclusion that K-12 educational tests are good indicators of student achievement and learning.

Although we claim no substantial expertise in how to resolve the unfortunate situation with the *Uniform Guidelines*, we offer some thoughts. We suggest that any reform in employment regulations be guided by scientific knowledge and professional practice. Thus, for example, all federal employment regulations should be fully consistent with the *Standards and Principles*. In addition, mechanisms should be established such that regulators rely on scientific knowledge as the basis for periodic revisions of regulations. Employment regulations would certainly benefit from scientific input. We call on regulatory agencies to issue an Advanced Notice of Proposed Rulemaking (ANPR). An ANPR issued for the *Uniform Guidelines* would be an invitation for public discussion on whether and how the

Uniform Guidelines need to be changed. Although we appreciate the role of attorneys in federal regulation, we assert that federal employment regulation will not improve until scientists, unaffiliated with the federal government, engage in a cooperative partnership with the regulatory process to alter the *Uniform Guidelines* so as to be consistent with science. We recommend that scientific organizations such as SIOP, partner with other professional organizations (e.g., Society of Human Resource Management, Equal Employment Advisory Council, Employment and Labor Law section of the American Bar Association) in promoting revisions to the regulations and in educating the Congress and the courts. What good is science if no one pays attention to it?

We encourage commentaries on this article to offer guidance concerning *how* the problems with the *Uniform Guidelines* can be remedied. That is, what are the reasonable next steps to cause federal regulation to be consistent with science? We also encourage commentaries on how Congress and the courts can be influenced to rely on scientific knowledge, even when the knowledge is politically and socially uncomfortable. Finally, given the emotive nature of this topic, we encourage collegial debate. With emotive topics, it is easy to offer opinions that yield more heat than light; it takes more work to consider the merits of both sides of an argument and to engage in a constructive, professional, and collegiate debate.

Science-Based Federal Regulations: A Role for SIOP

Unlike the agency authors of the *Uniform Guidelines*, many governmental agencies rely on science to form policy. For instance, the U.S. Food and Drug Administration's (FDA) mission depends on "science-led regulatory decisions" (FDA, 2011a). To ensure this, the FDA has 49 committees and panels to obtain expert advice on scientific, technical, and policy matters, including the Science Board to the FDA, whose role is to provide advice to FDA officials on scientific

and technical issues. Currently, all board members have doctorate degrees, and most are affiliated with major research universities (FDA, 2011b). The other committees and panels are associated with specific divisions within the FDA (Food, Drugs, Medical Devices, etc.). Although the scientific expertise is the top criterion in the selection process, other criteria such as potential conflict of interest are also evaluated (FDA, 2006). We acknowledge that federal regulation in employment testing does not likely need as many scientific advisory committees as the FDA, but scientific input into federal employment regulations is clearly warranted.

In addition to scientific panels guiding federal regulation, consumer advocacy organizations such as the Consumer Federation of America or the Center for Science in the Public Interest, both of which focus on nutrition and health and food safety, lobby for changes in laws and regulations. As an example of the successful intersection among lawmakers, advocacy organizations, and science, provisions in the Patient Protection and Affordable Care Act of 2010 require restaurants to display calorie information. It is likely that influence from consumer advocacy groups and scientific evidence (e.g., Burton, Creyer, Kees, & Huggins, 2006) has affected this law.

As another example of the intersection between science and federal regulations, several FDA guidelines specifically mention meta-analytic reviews as means to assess the efficacy of drugs. For instance, the FDA guidelines for the evaluation of cardiovascular risk in new antidiabetic therapies to treat type-2 diabetes (FDA, 2008) specifically state that meta-analyses of important cardiovascular events across clinical trials should be conducted. If federal employment regulation recognized meta-analysis as a form of validity documentation, the bad situation imposed on U.S. employers by federal employment regulators would be substantially improved.

We argue that the EEOC and related regulatory agencies could learn from the structure and processes used by the FDA.

In particular, a scientific advisory committee structure could guide the EEOC in the protection and advancement of equal employment opportunity laws and regulations. Currently, employment-related enforcement agencies appear to lack such an advisory committee structure. Certainly, such committees with independent experts would help to ensure that the regulatory process is transparent, which should increase the acceptance of science-led regulatory decisions by U.S. courts, Congress, businesses, employees, and the scientific communities.

SIOP's mission is to "enhance human well-being and performance in organizational and work settings by promoting the science, practice, and teaching of industrial and organizational psychology" (SIOP, n.d., p. A-1). Toward this end, SIOP has several objectives, including support of "SIOP members in their efforts to study, apply, and teach the principles, findings, and methods of industrial and organizational psychology"; the identification of "opportunities for expanding and developing the science and practice of industrial and organizational psychology"; the monitoring and addressing of "challenges to the understanding and practice of industrial and organizational psychology in organizational and work settings"; the promotion of "public awareness of the field of industrial and organizational psychology"; and the fostering of "cooperative relations with allied groups and professions" (SIOP, n.d., p. A-1).

Many of these objectives require the education of regulatory agencies, businesses, and the general public regarding the science and practice of I-O psychology. These objectives thus seem to call for an active role in the regulatory processes that affect scientists, practitioners, and businesses. To do this, SIOP has several committees, including the committee on Professional Practice, whose role is to "promote the interests of [SIOP] and its members by concerning itself with matters of professional practice and by developing relationships with other professional groups, business and

government leaders, and the public in general to advance the professional practice of industrial and organizational psychology" (SIOP, n.d., p. A-6). Other committees such as the Scientific Affairs and the State Affairs may also interact with external organizations, including federal and other regulatory agencies, to fulfill their roles.

Thus, SIOP's mission calls for, and its committee structure permits, the education of organizations including the employment regulatory agencies, the U.S. Congress, and U.S. courts. It is thus somewhat surprising that SIOP has not managed to build support from business and other organizations (e.g., the Society for Human Resource Management, the Equal Employment Advisory Council, and the Employment and Labor Law section of the American Bar Association) to voice the concerns in the scientific and business communities regarding the *Uniform Guidelines*. SIOP's inaction is counter to its mission. To fulfill its mission and maintain its scientific credibility, we recommend that SIOP become more proactive and involved in regulatory decision-making processes, new U.S. employment laws, and U.S. court decisions.

References

- Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology, 95*, 648–680. doi:10.1037/a0018714.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing* (2nd ed.). Washington, DC: American Educational Research Association.
- Banks, G. C., & McDaniel, M. A. (in press). Meta-analyses and selection procedures. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection*. Oxford, UK: Oxford University Press.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–23. doi:10.1111/j.1744-6570.1991.tb00688.x.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*, 9–30. doi:10.1111/1468-2389.00160.

- Bartlett, C. J., Bobko, P., Mosier, S. B., & Hannan, R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. *Personnel Psychology, 31*, 233–241. doi:10.1111/j.1744-6570.1978.tb00442.x.
- Biddle, D. A. (2010). Should employers rely on local validation studies or validity generalization (VG) to support the use of employment tests in Title VII Situations? *Public Personnel Management, 39*, 307–326.
- Boehm, V. R. (1977). Differential prediction: A methodological artifact? *Journal of Applied Psychology, 62*, 146–154. doi:10.1037/0021-9010.62.2.146.
- Borneman, M. J. (2010). Using meta-analysis to increase power in differential prediction analyses. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3*, 224–227. doi:10.1111/j.1754-9434.2010.01228.x.
- Bray, D. W., & Moses, J. L. (1972). Personnel selection. *Annual Review of Psychology, 23*, 545–576. doi:10.1146/annurev.ps.23.020172.002553.
- Burton, S., Creyer, E. H., Kees, J., & Huggins, K. (2006). Attacking the obesity epidemic: The potential health benefits of providing nutrition information in restaurants. *American Journal of Public Health, 96*, 1669–1675. doi:10.2105/AJPH.2004.054973.
- Cascio, W. E., & Aguinis, H. (2001). The federal uniform guidelines on employee selection procedures (1978). An update on selected issues. *Review of Public Personnel Administration, 21*, 200. doi:10.1177/0734371X0102100303.
- Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap glosing: When the “have-nots” gain but the “haves” gain even more. *American Psychologist, 60*, 149–160. doi:10.1037/0003-066x.60.2.149.
- Cohen, M. S., Aamodt, M. G., & Dunleavy, E. M. (2010). *Technical advisory committee report on best practices in adverse impact analyses*. Washington, DC: Center for Corporate Equality.
- Copus, D. A. (2006). *Validation of cognitive ability tests*. Letter to Charles James, Office of Federal Contract Compliance Programs (March 27, 2006). Morristown, NJ: Ogletree Deakins.
- Daniel, C. (2001). Separating law and professional practice from politics. *Review of Public Personnel Administration, 21*, 175. doi:10.1177/0734371X0102100301.
- Equal Employment Opportunity Commission (EEOC). (1966). Guidelines on employment testing procedures. *Federal Register, 31*, 6414.
- Equal Employment Opportunity Commission (EEOC). (1970). Guidelines on employee selection procedures. *Federal Register, 35*, 12333–12336.
- Equal Employment Opportunity Commission (EEOC), Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register, 43*, 38290–39315.
- Ewoh, A. I. E., & Guseh, J. S. (2001). The status of the uniform guidelines on employee selection procedures. *Review of Public Personnel Administration, 21*, 185. doi:10.1177/0734371X0102100302.
- Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology, 61*, 579–616. doi:10.1111/j.1744-6570.2008.00123.x.
- Food and Drug Administration (FDA). (2006). *FDA announces plan to strengthen advisory committee processes*. Retrieved from <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2006/ucm108697.htm>.
- Food and Drug Administration (FDA). (2008). *Guidance for industry: Diabetes mellitus—Evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes*: U.S. Food and Drug Administration.
- Food and Drug Administration (FDA). (2011a). *About science & research at FDA*. Retrieved from <http://www.fda.gov/ScienceResearch/AboutScience/ResearchatFDA/default.htm>.
- Food and Drug Administration (FDA). (2011b). *Science board to the food and drug administration*. Retrieved from <http://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/ScienceBoardtotheFoodandDrugAdministration/default.htm>.
- Gatewood, R. D., Feild, H. S., & Barrick, M. R. (2008). *Human resource selection* (6th ed.). Mason, OH: South-Western.
- Grutter v. Bollinger. (2003). 539 U.S. 306.
- Guion, R. M. (1965). *Personnel testing*. New York, NY: McGraw-Hill.
- Guion, R. M. (1975). Recruitment, selection and job placement. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago, IL: Rand McNally.
- Hartigan, J. A., & Wigdor, A. K. (Eds.). (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery*. Washington, DC: National Academy Press.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York, NY: Free Press.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72–98. doi:10.1037/0033-2909.96.1.72.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. (2nd ed.). Newbury Park, CA: Sage.
- Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin, 86*, 721–735. doi:10.1037/0033-2909.86.4.721.
- Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91*, 594–612. doi:10.1037/0021-9010.91.3.594.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869–879. doi:10.1037/0021-9010.85.6.869.
- Jeanneret, P. R. (2005). Professional and technical authorities and guidelines. In F. J. Landy (Ed.), *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives* (pp. 47–100). San Francisco, CA: Wiley.

- Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.
- Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1–123.
- Kirkpatrick, J. J., Ewen, R. B., Barrett, R. S., & Katzell, R. A. (1968). *Testing and fair employment*. New York, NY: New York University Press.
- Kleiman, L. S., & Faley, R. H. (1985). The implications of professional and legal guidelines for court decisions involving criterion-related validity: A review and analysis. *Personnel Psychology*, 38, 803–833. doi:10.1111/j.1744-6570.1985.tb00568.x.
- Kravitz, D. A. (2008). The diversity-validity dilemma: Beyond selection—the role of affirmative action. *Personnel Psychology*, 61, 173–193. doi:10.1111/j.1744-6570.2008.00110.x.
- McDaniel, M. A. (2007). Validity generalization as a test validation approach. In S. M. McPhail (Ed.), *Alternative validation strategies: Developing new and leveraging existing validity evidence*. (pp. 159–180). Hoboken, NJ: Wiley.
- McDaniel, M. A. (2009). Gerrymandering in personnel selection: A review of practice. *Human Resource Management Review*, 19, 263–270. doi:10.1016/j.hrmr.2009.03.004.
- McDaniel, M. A. (2010, July). *Abolish the uniform guidelines*. Paper presented at the annual meeting of the International Personnel Assessment Council, Newport Beach, CA.
- McKay, P. F., & McDaniel, M. A. (2006). A reexamination of Black-White mean differences in work performance: More data, more moderators. *Journal of Applied Psychology*, 91, 538–554. doi:10.1037/0021-9010.91.3.538.
- Meade, A. W., & Tonidandel, S. (2010). Not seeing clearly with Cleary: What test bias analyses do and do not tell us. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3, 192–205. doi:10.1111/j.1754-9434.2010.01223.x.
- Michaels, D., & Monforton, C. (2005). Manufacturing uncertainty: Contested science and the protection of the public's health and environment. *American Journal of Public Health*, 95, S39–S45. doi:10.2105/AJPH.2004.043059.
- Morris, S. B., & Lobsenz, R. E. (2000). Significance tests and confidence intervals for the adverse impact ratio. *Personnel Psychology*, 53, 89–111. doi:10.1111/j.1744-6570.2000.tb00195.x.
- National Center for Education Statistics. (2010). *Digest of education statistics; Table 386. Literacy skills of adults, by type of literacy, proficiency levels, and selected characteristics: 1992 and 2003*. Retrieved from http://nces.ed.gov/programs/digest/d09/tables/dt09_386.asp.
- O'Boyle, E. H., & McDaniel, M. A. (2008). Criticisms of employment testing: A commentary. In R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing*. (pp. 181–197). Washington, DC: American Psychological Association.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65, 373–406. doi:10.1037/0021-9010.65.4.373.
- Phillips, M., Brooks-Gunn, J., Duncan, G. J., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap*. Washington, DC: Brookings Institution Press.
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153–172. doi:10.1111/j.1744-6570.2008.00109.x.
- Pyburn, K. M., Jr., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology*, 61, 143–151. doi:10.1111/j.1744-6570.2008.00108.x.
- Rosenstock, L., & Lee, L. (2002). Attacks on science: The risks to evidence-based policy. *American Journal of Public Health*, 92, 14–18. doi:10.2105/AJPH.92.1.14.
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, 54, 297–330. doi:10.1111/j.1744-6570.2001.tb00094.x.
- Roth, P. L., Bobko, P., & Switzer, F. S., III. (2006). Modeling the behavior of the 4/5ths rule for determining adverse impact: Reasons for caution. *Journal of Applied Psychology*, 91, 507–522. doi:10.1037/0021-9010.91.3.507.
- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, 56, 302–318. doi:10.1037/0003-066x.56.4.302.
- Schenkel, R. (2010). The challenge of feeding scientific advice into policy-making. *Science*, 330, 1749–1751. doi:10.1126/science.1197503.
- Schmidt, F. L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior*, 33, 272–292. doi:10.1016/0001-8791(88)90040-1.
- Schmidt, F. L., Gast-Rosenberg, I., & Hunter, J. E. (1980). Validity generalization results for computer programmers. *Journal of Applied Psychology*, 65, 643–661. doi:10.1037/0021-9010.65.6.643.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529–540. doi:10.1037/0021-9010.62.5.529.
- Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128–1137. doi:10.1037/0003-066x.36.10.1128.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274. doi:10.1037/0033-2909.124.2.262.
- Schmidt, F. L., & Hunter, J. E. (2003). History, development, evolution, and impact of validity generalization and meta-analysis methods, 1975–2001. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 31–65). Mahwah, NJ: Erlbaum.

- Schmidt, F. L., Pearlman, K., & Hunter, J. E. (1980). The validity and fairness of employment and educational tests for Hispanic Americans: A review and analysis. *Personnel Psychology, 33*, 705–724. doi: 10.1111/j.1744-6570.1980.tb02364.x.
- Schmitt, N., & Quinn, A. (2010). Reductions in measured subgroup mean differences: What is possible? In J. L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 425–451). New York, NY: Routledge.
- Sharf, J. (2006). *Letter to Cari M. Dominguez, Chair, Equal Employment Opportunity Commission (May 10, 2006)*. Alexandria, VA: Author.
- Sharf, J. (2008, February). *Enforcement agencies' response to validity generalization*. Paper presented at the annual meeting of the Personnel Testing Council of Metropolitan Washington, Washington, DC.
- Shoben, E. W. (1978). Differential pass-fail rates in employment testing: Statistical proof under Title VII. *Harvard Law Review, 91*, 793–813.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*, 15–21. doi: 10.3102/0013189X031007015.
- Society for Industrial and Organizational Psychology (SIOP). (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.
- Society for Industrial and Organizational Psychology (SIOP). (n.d.). *SIOP bylaws*. Retrieved from <http://www.siop.org/reportsandminutes/bylaws.pdf>.
- Steinbrook, R. (2004). Peer review and federal regulations. *New England Journal of Medicine, 350*, 103–104. doi:10.1056/NEJMp038230.
- Stillwell, R. (2010). *Public school graduates and dropouts from the common core of data: School year 2007-08. NCES 2010-341*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- U.S. Census Bureau. (2007). *Latest SUSB annual data: U.S. & states, totals*. Retrieved from <http://www.census.gov/econ/susb/>.
- Walberg, H. J., & Tsai, S.-L. (1983). Matthew effects in education. *Educational Research Quarterly, 20*, 359–373. doi:10.2307/1162605.
- Wigdor, A. K., & Garner, W. R. (Eds.). (1982). *Ability testing: Use, consequences, and controversies*. Washington, DC: National Academy Press.