Are the Uniform Guidelines Outdated?

Federal Guidelines, Professional Standards, and Validity Generalization (VG)

Daniel A. Biddle, Ph.D.

CEO, Biddle Consulting Group, Inc.
Abstract

In the 1971 Griggs v. Duke Power case, 8 Supreme Court justices unanimously agreed that a test with adverse impact is unlawfully discriminatory unless it has been validated in accord with the Guidelines (Guidelines, Q&A #2). This ruling has governed EEO enforcement for the past 37 years. Since that time, several versions of the professional standards (SIOP Principles and the Joint Standards) have been published, and the most recent of which endorse Validity Generalization (VG) studies as one of the five sources of validity evidence. However, employers that use selection procedures that cause adverse impact are required by the Uniform Guidelines (based on Griggs) to make a local demonstration of validity—a requirement that creates a challenge for I-O practitioners.
The Uniform Guidelines on Employee Selection Procedures

Two years after the 1964 Civil Rights Act passed into law, the Equal Employment Opportunity Commission (EEOC) published the first set of Guidelines relevant to employment testing—the Guidelines on Employment Testing Procedures (August 24, 1966). These Guidelines interpreted “professional ability tests” to mean: “a test which fairly measures the knowledge or skills required by the particular job or class of jobs which the applicant seeks, or which fairly affords the employer a chance to measure the applicant’s ability to perform a particular job or class of jobs.” The EEOC published another version four years later titled, “Guidelines on Employee Selection Procedures” (August 1, 1970). One year later, the U.S. Supreme Court tried the first post-Civil Rights Act case: Griggs v. Duke Power (1971). The unanimous Griggs decision held that selection procedures—including tests—that caused adverse impact had to be justified through a process of demonstrating job relatedness and business necessity.

Seven years after Griggs, four federal agencies (the U.S. Department of Justice, Department of Labor, the Equal Employment Opportunity Commission, and the Civil Service Board) released an updated version of the federal Guidelines—today simply known as the Uniform Guidelines on Employee Selection Procedures (1978). Their stated purpose for framing the Guidelines was to:

. . . incorporate a single set of principles which are designed to assist employers, labor organizations, employment agencies, and licensing and certification boards to comply with requirements of Federal law prohibiting employment practices
which discriminate on grounds of race, color, religion, sex, and national origin (Guidelines, Section B).

The Guidelines were thus published to provide guidance on how to comply with federal law—i.e., the recent Griggs case. The Guidelines continue to state that failure to validate a selection procedure in accordance with their criteria therein constitutes discrimination for the employer using the procedure (Guidelines, Section 3A; Questions & Answers, #2).

This document has since been used in thousands of government enforcement and judicial settings where employers have been required to demonstrate that their selection procedures causing adverse impact are sufficiently “job related” by addressing the Guidelines requirements. A companion “Questions & Answers” document was finalized on May 2, 1980 that included 93 questions and answers regarding some of the topics covered by the Guidelines. There have been no additions or changes since.


Industrial-organizational (I-O) psychologists are not unaware of the professional standards, consisting of the Joint Standards (1999) and SIOP Principles (2003). The stated purpose of the Joint Standards is to “. . . provide criteria for the evaluation of tests, testing practices, and test use” for professional test developers, sponsors, publishers, and users that adopt Standards (p. 2). One of the fifteen chapters (Chapter 14) is devoted
exclusively to testing in the areas of employment and credentialing. The remaining chapters pertain to developing, administering, and using tests of various sorts. An updated version is expected in 2008-2009.

SIOP released the updated Principles in 2003, which are designed “to a large degree a technical document, but it is also an informational document.” While covering many of the same topics included in the Guidelines, the Principles include a caveat with respect to the legal aspects surrounding testing: “Federal, state, and local statutes, regulations, and case law regarding employment decisions exist. The Principles is not intended to interpret these statutes, regulations, and case law, but can inform decision making related to them” (p. 1). The Joint Standards hold a similar status.

This constitutes a major distinction between the Guidelines and the professional standards. The Guidelines only apply whenever an employer’s selection procedure has adverse impact. The professional standards embody best practice guidelines that apply to situations where adverse impact may or may not exist.

When the Guidelines were published, the professional standards were current as of 1974 (Standards) and 1975 (Principles). Fortunately, the framers of the Guidelines had sufficient foresight that future editions of these standards would be updated to reflect innovations in the field of measurement theory, so a sort of auto-updating clause was added:

Question: What is the relationship between the validation provisions of the Guidelines and other statements of psychological principles, such as the Standards for Educational and Psychological Tests, published by the

_Are the Uniform Guidelines Outdated? 5_
Work in Progress • Biddle Consulting Group, Inc.
American Psychological Association…? Answer: The validation provisions of the Guidelines are designed to be consistent with the generally accepted standards of the psychological profession. *These Guidelines also interpret Federal equal employment opportunity law,* and embody some policy determinations of an administrative nature. To the extent that there may be differences between particular provisions of the Guidelines and expressions of validation principles found elsewhere, the Guidelines will be given precedence by the enforcement agencies. (Guidelines, Q&A #40). (emphasis added)

The Guidelines’ deference to legal requirements (rather than professional standards) has also been observed in litigation settings. For example, in *Lanning v. SEPTA* (1999), the 3rd Circuit Court of Appeals stated: “To the extent that the SIOP Principles are inconsistent with the mission of *Griggs* and the business necessity standard adopted by the Act, they are not instructive” (FN20).¹

Because of these distinctions, the Guidelines have been cited hundreds of times in state and federal Title VII cases. By contrast, the professional standards have been collectively cited fewer than 40 times (based on Westlaw searches as of the date of this writing). The Guidelines have also been cited as the sole standard for validity review in numerous cases (e.g., a Westlaw search for “validated in accordance with the guidelines” returned 44 distinct cases).

¹ U.S. v. City of Erie (PA 411 F.Supp.2d 524 W.D. Pa., 2005, FN 18) clarified this criticism stating that the Lanning court did not “throw out” or otherwise invalidate the SIOP Principles in their entirety when making this statement.
One of the most notable distinctions between the Guidelines and professional standards is the coverage and intended audience of each. The Guidelines are written expressly to employers that are subject to Title VII. They are utilized by employers and federal enforcement agencies to evaluate validity when an employer’s testing practices have adverse impact. The professional standards, however, are written primarily to and for professionals in the test development field and constitute a set of technical standards for developing and evaluating tests.

The Guidelines and professional standards also differ with respect to their fundamental purpose. The stated purpose of the Guidelines is to help employers comply with requirements of Federal law prohibiting employment practices which discriminate on grounds of race, color, religion, sex, and national origin and to “. . . provide a framework for determining the proper use of tests and other selection procedures” (Section 1B). The stated purpose of the professional standards is to provide criteria for the evaluation of tests, testing practices, and test use for professional test developers, sponsors, publishers, and users that adopt the Standards and to “address the needs of persons involved in personnel selection” (Principles, p. 1).

These distinctions are not without practical impact to the I-O practitioner. Consider a fundamental concept at the heart of all test validity research: test reliability. Both sets of the most recent professional standards have extensive coverage of this critical topic. In fact, the Joint Standards dedicate an entire chapter to this topic which reviews important developments that have come to the forefront of reliability theory in recent years (e.g., using conditional standard errors of measurement and evaluating the “decision consistency” reliability of tests). The Guidelines do not even define test
reliability—they simply state that *having* reliability is essential when employers are making a validity defense. The Guidelines require users to report reliability, but do not provide any application guidelines such as those provided by the professional standards. By contrast, the Guidelines exclusively remark about “job relatedness” and “business necessity” requirements, adverse impact, and other matters relevant to Title VII.

Title VII, the Guidelines, Professional Standards, and Validity Generalization (VG)

One of the contentious topics that sometimes emerge in I-O forums is Validity Generalization (VG). VG studies combine the results of statistical validation studies to evaluate the effectiveness (i.e., validity) of a personnel test or particular type of tests and to describe what the findings mean in a *broader, more general sense* (Murphy, 2003). The mission and objective of the Guidelines differ from VG. Because the Guidelines come into force whenever an employer’s *particular testing practice* has adverse impact, they are concerned with *validity specificity*. To wit, they are narrowly-targeted to assist federal enforcement agencies to answer this question: Is this particular employer’s testing practice “job related *for the position in question* and consistent with business necessity” (1991 Civil Rights Act, 2000e-2[k][1][A][i]). Thus, the Guidelines are narrowly-tailored to evaluate whether a *specific test* is sufficiently valid for a *specific position*.

VG is tailored at answering a different question: “How broadly does this test (or construct) correlate to job performance across different positions, settings, and employers? This is a fundamentally different question than the one targeted by the
Guidelines. Even if a test used by an employer “shows up valid” for 100 other positions/employers, the challenged employer still has the burden for showing the test is job related for their position in question and consistent with business necessity in their context.

When it comes to employers relying on statistical validity evidence from other positions/employers, the Guidelines require that employers to conduct a transportability study. Specifically, Section 7B of the Guidelines require that the “other” validation studies sufficiently meet the Guidelines requirement, that the jobs in the other validation studies are highly similar to the target position, and evidence is provided that the test is a fair predictor of job performance (i.e., that it isn’t biased against certain groups).

The professional standards also provide standards for adopting validity evidence from outside situations. However, even though the professional standards permit borrowing validity evidence, employers should proceed with caution when doing so because some have been met with harsh criticism in the courts when attempting to use VG to generalize validity into their setting.

For example, the Sixth Circuit Court of Appeals ruled that VG, as a matter of Title VII law, could not be used to justify Atlas Paper’s testing practices that had adverse impact (EEOC v. Atlas Paper, 1989). In Atlas, the Sixth Circuit rejected the use of VG to justify a test purporting to measure general intelligence which had adverse impact when used for screening clerical employees. Without conducting a local validity study, an expert testified regarding the generalized validity of the test, stating that it was “valid for all clerical jobs.” The lower District Court had previously approved Atlas’ use of the test,
but the Court of Appeals reversed this decision and rejected the use of VG evidence as a basis for justifying the use of the test by stating:

We note in respect to a remand in this case that the expert failed to visit and inspect the Atlas office and never studied the nature and content of the Atlas clerical and office jobs involved. The validity of the generalization theory utilized by Atlas with respect to this expert testimony under these circumstances is not appropriate. Linkage or similarity of jobs in dispute in this case must be shown by such on site investigation to justify application of such a theory.

The criteria applied by the court in this case is exactly what is required by the Guidelines for transporting validity evidence into a new situation (Section 7B)—i.e., conducting a job comparability study. Even the authors who published the genesis VG article in the field of personnel selection originally advocated the job process be conducted for transporting validity evidence (Schmidt & Hunter, 1977, p. 530). The Sixth Circuit decision in Atlas continued to offer a more direct critique of VG by stating:

The premise of the validity generalization theory, as advocated by Atlas’ expert, is that intelligence tests are always valid. The first major problem with a validity generalization approach is that it is radically at odds with Albemarle Paper v. Moody, Griggs v. Duke Power, relevant case law within this circuit, and the EEOC Guidelines, all of which require a
showing that a test is actually predictive of performance at a specific job. The validity generalization approach simply dispenses with that similarity or manifest relationship requirement. Albemarle and Griggs are particularly important precedents since each of them involved the Wonderlic Test . . . Thus, the Supreme Court concluded that specific findings relating to the validity of one test cannot be generalized from that of others” (EEOC v. Atlas Paper, 868 F.2d. at 1499).

The judge issued a factual conclusion based upon the applicability of the U.S. Supreme Court Albemarle (1975) case findings regarding the situational specific validity requirements and made a factual conclusion and rule of law, stating:

The kind of potentially Kafkaesque result, which would occur if intelligence tests were always assumed to be valid, was discussed in Van Aken v. Young (451 F.Supp. 448, 454, E.D. Mich. 1982, aff’d 750 F.2d. 43, 6th Cir. 1984). These potential absurdities were exactly what the Supreme Court in Griggs and Albemarle sought to avoid by requiring a detailed job analysis in validation studies. As a matter law . . . validity generalization theory is totally unacceptable under the relevant case law and professional standards (EEOC v. Atlas Paper, 868 F.2d. at 1499).

Summary
Since the federal Guidelines were enacted, the I-O community has seen the release of three updated versions of the Principles (1980, 1987, and 2003) and two updated versions of the Joint Standards (1985 and 1999). With these updates, one is left to wonder if the Guidelines need to follow suit and be revised. However, since the Guidelines are essentially based on the federal Civil Rights Act and cornerstone U.S. Supreme Court cases such as Griggs and Albemarle, one must first ask, “Is the Civil Rights Act outdated? Are the Griggs and Albemarle cases outdated?”

The Griggs case started a chain of events that have created lasting foundations in the field of EEO enforcement. The legal principles laid down in Griggs were endorsed by Congress, continually reaffirmed by the Supreme Court, and were then incorporated into the Guidelines. After the Guidelines were published, the job relatedness burden defined by Griggs and interpreted by the Guidelines has been subsequently endorsed in thousands of government enforcement and legal settings.

Concurrent with these developments, technical innovations in testing and measurement theory have continued to evolve and marked with periodic updates to the professional standards. However, the legal requirement of “demonstrating job relatedness for the position in question” originally provided by Griggs and subsequently interpreted by the Guidelines remain intact. This is where the Griggs-Guidelines requirement seem to diverge from the goal of VG as defined by the professional standards. While Title VII requires the employer to demonstrate that their selection procedure is job related for the position in question, VG studies are conducted to evaluate how the validity of a particular test or construct may generalize across settings and positions.
If the first burden in Title VII settings (proving adverse impact) cannot be carried using evidence solely from external locations, it seems to follow that the second burden (proving validity) should likewise not be provable using only external evidence. One can only imagine the outcry of defense attorneys if government enforcement agencies or plaintiff attorneys were permitted to transport or generalize adverse impact into a local employer based on adverse impact that occurred “at some other location.”

VG can be an important tool for identifying the selection procedures that might be appropriate for use by employers. However, the federal Guidelines require that a study of the similarity of the target job in question and the jobs for which the selection procedure had been previously found to be valid should be conducted. This approach embodies both the letter and spirit of nondiscrimination. By conducting such a similarity study the employer would be testing the appropriateness of a particular selection procedure as it is used for a particular job by a particular employer. This seems a reasonable demand because employers following this process will likely benefit from both increased defensibility as well as increased utility when selection procedures are carefully matched to the requirements of the target position (Dye, Reck, & McDaniel, 1993).
References

1991 Civil Rights Act (42 U.S.C. §2000e-2[k][1][A][i]).


