Abolish the Uniform Guidelines

Michael A. McDaniel  
Virginia Commonwealth University  
Invited Address to IPAC 2010

Biographical Introduction

- I am an Industrial & Organizational psychologist.
- I have worked in personnel selection since the late 1970s, both in operational staffing positions and in research/academic positions.
- Most of my research concerns personnel selection and factors that influence adverse impact.

“Uniform Guidelines”

- Uniform Guidelines on Employee Selection Procedures (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor and Department of Justice, 1978)
- Given deference in employment litigation.
- The Uniform Guidelines are 32 years old.
Goals of Presentation

• My presentation is best described as *advocacy* in that I call for the abolishment, or at least a major revision, of the Uniform Guidelines.
• My arguments are primarily based on scientific research and the inconsistency of the Uniform Guidelines with scientific knowledge and professional practice.

Acknowledgement

• I would like to thank Dr. Richard F. Tonowski, Chief Psychologist at the Equal Employment Opportunity Commission.
• One of his comments motivated me to begin arguing for the abolishment of the Uniform Guidelines.
• Alleged to have said the Uniform Guidelines will outlive him.
  - 20 to 30 more years

Acknowledgement

• In my opinion, the Uniform Guidelines are sufficiently flawed, such that I cannot stomach the possibility that they would be influencing personnel selection for several more decades.
• Thus, I offer this paper.
Raise the level of debate

• Currently no one talks much about abolishing the Uniform Guidelines, despite their inconsistency with scientific knowledge and professional practice.
• I seek more people to talk about the possibility of the abolishment or extensive revision of the Uniform Guidelines.

Raise the level of debate

• I am hoping that professional organizations concerning personnel selection (e.g., IPAC, PTC’s, SIOP) will begin a debate concerning the abolishment or extensive revision of the Uniform Guidelines.
• I am hoping they begin advocacy efforts to support the abolishment or extensive revision of the Uniform Guidelines.

Raise the level of debate

• By advocacy efforts, I refer to actions such as:
  - Issuing policy statements concerning the flaws of the Uniform Guidelines
  - Contacting the organizational authors of the Uniform Guidelines
  - Educating our elected Federal congressional representatives and senators
  - Educating attorneys and federal judges
Raise the level of debate

- I don’t claim to have all the answers for the problems associated with the Uniform Guidelines.
- I will share what I think about the problems in the Uniform Guidelines and hope that my talk will encourage constructive debate on what to do concerning the flaws of the Uniform Guidelines.

What I am not arguing:

- I am not arguing anything counter to:
  - Civil Rights Act
  - Equal employment opportunity
  - Affirmative action and diversity efforts
  - Good professional practice
  - Continued research in understanding and reducing sub-group differences
  - Continued research in personnel selection

Disclosures

- When the Uniform Guidelines were issued in 1978, they were sometimes called the full employment act for industrial and organizational psychologists.
- The Uniform Guidelines influenced employers to hire people with my training.
- The Uniform Guidelines also resulted in me getting consulting and expert witness work.
Disclosures

• Such work permitted me to pay down my mortgage and save for my children’s college education.
• Despite the financial and job benefits that I have received as a result of the Uniform Guidelines, I will be arguing that the Uniform Guidelines are damaging to our discipline of personnel selection.

Questions and Expression of Other Viewpoints

• Not everyone in the room will accept my arguments.
• I will finish with enough time left for questions and opposing viewpoints.
• Again, my goal is to foster honest and open debate on the Uniform Guidelines.

Other Guidelines

• In part, I will be contrasting the Uniform Guidelines with two other guidelines intended to guide practice in personnel selection.
• These are not government issued guidelines, but guidelines offered by scientific professional organizations who based their recommendations on scientific knowledge and professional practice.
“Principles”


“Standards”


The 70’s

- Although this is a serious talk about a serious issue, I would like to offer a few other items from the 1970’s to give us hope that the Uniform Guidelines will not always be with us.
1970’s Kitchen

1970’s Hair Style

1970’s Shag Carpet
1970’s Roller Disco

Uniform Guidelines

• Unlike avocado green refrigerators, orange metallic wall paper, some bizarre hair styles, shaved shag rugs, roller disco, and other 1970’s phenomena that are fortunately seldom seen anymore, the Uniform Guidelines are still with us.

Inconsistent with scientific knowledge and professional practice
The Uniform Guidelines Unfulfilled Promises

• The Uniform Guidelines stated that they are intended to be consistent with scientific knowledge and professional practice.
• This has not happened.

60-3.5 - General standards for validity studies.

A. ...New strategies for showing the validity of selection procedures will be evaluated as they become accepted by the psychological profession.

C. Guidelines are consistent with professional standards. 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 [...] and standard textbooks and journals in the field of personnel selection.
Unfulfilled Commitments

• Despite the commitments made in the Uniform Guidelines, the Uniform Guidelines have never been updated in 32 years.
• Contrast this behavior with other Federal regulators like the Food and Drug Administration (FDA). The FDA updates its regulations and drug approvals as new science becomes available.

Thus, the following organizational authors of the Uniform Guidelines are not doing their job:
- Equal Employment Opportunity Commission
- Office of Personnel Management (formerly Civil Service Commission)
- Department of Labor
- Department of Justice

The 1970’s Perspective: Situational Specificity

• Beginning in the 1920’s and continuing into the 1970s, it was observed that the same employment test yielded different validity results across settings.
• For example, a test to screen bank clerks might have a high validity (e.g., high correlation with job performance) in one bank but not in the bank across the street.
The 1970's Perspective: Situational Specificity

- This led to the speculation that there were yet to be uncovered characteristics of the employment situation that caused an employment test to be valid in one situation (e.g., one bank) but not valid in another situation (e.g., the bank across the street).
- This speculation became known as the situational specificity hypothesis.

The 1970's Perspective: Situational Specificity

- A large emphasis was placed on detailed job analyses in hope of finding the causes of situational specificity.
- However, the job analyses, despite their detail and comprehensiveness, never identified the characteristics that caused a test to be valid in one situation but not in another.

The 1970's Perspective: Situational Specificity

- Because test validity appeared to vary across situations and no one knew why, many argued that one should conduct a validity study in each application of the test.
  - Local validity studies
  - The bank on one side of the street and the bank across the street both need to do validity studies on the same test.
Refutation of the situation specificity hypothesis

- Beginning in 1977, Schmidt and Hunter began offering data to demonstrate that the situational specificity hypothesis was false.
- They demonstrated that random sampling error was responsible for most of the variability in validities across studies.

Refutation of the situation specificity hypothesis: Sampling error

- Sampling error: Consider a bag of 100 poker chips in which some are red and the remaining are white.
- One seeks to infer the proportion of red poker chips by drawing a sample of 10 poker chips from the bag.

- Manny draws 10 chips of which 8 are red.
- Moe draws 10 chips of which 6 are red.
- Jack draws 10 chips of which 7 are red.
- The percentage of red poker chips in the bag is constant, yet the samples yield different results.
- Different samples drawn from the bag yield different results due to random sampling error.
Refutation of the situation specificity hypothesis: Sampling error

- When there is less than an infinite number of observations (people or poker chips) in the sample, the results of the various samples will not agree due to sampling error.

Thus, even if the validity of a test for bank clerks is constant in the population of all bank clerks, one will get different validity estimates from different banks. Thus the apparent situational specificity of employment test validity is actually the result of random sampling error from small samples.

In the 1970s, the average sample size in a validity study was 68 people. We now know that this sample size is too small to estimate accurately the validity of a test. Consider a test with a correlation of .20 and we draw samples of 68 people. The sample validity estimates would range from -.04 to .42.
Refutation of the situation specificity hypothesis: Sampling error

- The small sample sizes, with large amounts of sampling error, made validities appear to be unstable even though they were often very stable in the population.

Uniform Guidelines Flaw #1: Emphasis on local validation studies

- Unfortunately, the Uniform Guidelines had been published before the refutation of the situational specificity hypothesis became well known.
- As a result, the Uniform Guidelines emphasize local validity studies.

Uniform Guidelines Flaw #1: Emphasis on local validation studies

- In contrast to the Uniform Guidelines, which were never updated to reflect changes in scientific knowledge, the Principles and the Standards recognize the futility of local validation studies without large sample sizes.
Meta-analysis

• The work of Schmidt & Hunter in the late 1970s refuting situational validity evolved into methods to cumulate validity evidence across studies.
• The resulting methods, now known as meta-analysis, cumulate research results across studies.
• The methods are used in many disciplines.
  - Meta-analysis is the basis of evidence-based medicine.

Meta-analysis

• Initially this Schmidt & Hunter meta-analytic research stream was called validity generalization because it demonstrated that most employment tests were valid across situations (i.e., the validity generalized).
  - Largely eliminating the need for local validation

Uniform Guidelines Flaw #2: Ignoring Meta-Analytic Research

• The Uniform Guidelines have never been revised to acknowledge the very useful role of meta-analysis in demonstrating validity of employment tests.
• The Principles and Standards recognize the extensive research concerning meta-analysis and offer it as a useful approach to demonstrating validity.
Uniform Guidelines Flaw #3: Job Analysis – My Way or the Highway

- Although detailed and comprehensive job analyses have seldom found any situational characteristics that influence validity, in 1978 there was still hope that this could be accomplished.
- Thus, the authors of the Uniform Guidelines incorporated the requirement for a detailed job analysis as part of a content validity defense.

Uniform Guidelines Flaw #3: Job Analysis – My Way or the Highway

- The Uniform Guidelines endorsed a specific approach to job analysis for a content validity defense. As typically implemented, it is very detailed, very labor intensive, and thus costly.
- There is no scientific evidence that the approach required in the Uniform Guidelines is superior to any other job analysis method for establishing content validity.

Uniform Guidelines Flaw #3: Job Analysis – My Way or the Highway

- In contrast to the Uniform Guidelines, which have never been revised to reflect the diverse ways of demonstrating content validity, the Standards and Principles rely on scientific knowledge and endorse a variety of approaches as evidence of validity.
Uniform Guidelines Flaw #4: The Assassination of Construct Validity

• The Uniform Guidelines assassinated construct validity by defining it in a way such that no one will want to use it.

Quotes from D. Technical standards for construct validity studies

- “Construct validity is a more complex strategy”
- “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.”

Unlike the Uniform Guidelines which have not been revised with respect to scientific or professional practice, the Principles and Standards recognize the value of construct evidence in demonstrating the validity of tests.
Uniform Guidelines Flaw #5: 1950's Perspective on Separate Validity Methods

• The Uniform Guidelines view three distinct methods for validity documentation, these being:
  - Content validity
  - Construct validity
  - Criterion validity
• This approach was conceptualized in the 1950s.

Uniform Guidelines Flaw #5: 1950’s Perspective on Separate Validity Methods

• The Uniform Guidelines have specific rules or approaches that must be followed for each “validity method.”
• There is no scientific evidence supporting the superiority of the Uniform Guidelines approach relative to other approaches.
• “There is no single approach that is the preferred method for the analysis of work.” Principles, p. 11

Uniform Guidelines Flaw #5: 1950’s Perspective on Separate Validity Methods

• Professional practice as codified in the Standards and Principles, now view “validity as a unitary concept with different sources of evidence contributing to an understanding of the inferences that can be drawn from a selection procedure.” (Principles, p 4)
Uniform Guidelines Flaw #5: 1950’s Perspective on Separate Validity Methods

- “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.” Principles, p. 4

Uniform Guidelines Flaw #5: 1950’s Perspective on Separate Validity Methods

- “The Standards discusses five sources of evidence that can be used in evaluating a proposed interpretation of selection procedure test scores for a particular use:
  - (a) relationships between predictor scores and other variables, such as test-criterion relationships,
  - (b) content,
  - (c) internal structure of the test,
  - (d) response processes, and
  - (e) consequences of testing.” Principles p 5.

Uniform Guidelines Flaw #5: 1950’s Perspective on Separate Validity Methods

- “Given that validity is a unitary concept, such categorizations refer to various sources of evidence rather than distinct types of validity.
- It is not the case that each of these five sources is an alternative approach to establishing job relatedness.
- Rather, each provides information that may be highly relevant to some proposed interpretations of scores, and less relevant, or even irrelevant to others.” Principles, p 5.
Uniform Guidelines Flaw #5: 1950's Perspective on Separate Validity Methods

- The Uniform Guidelines have never been revised to bring their 1950's conceptualization of validity into compliance with the Standards and the Principles.

Adverse Impact

- Adverse impact refers to differences in hiring rates by demographic sub-groups.
- The Uniform Guidelines include a decision rule called the 4/5ths rule to determine if adverse impact is present.
- This is a rule of thumb. It has no basis in science.

Adverse Impact

- More recently, EEOC and the Office of Federal Contract Compliance Programs (OFCCP) are using a two-standard deviation test that was originally introduced into case law in a voting rights case.
- Because this two-standard deviation test is a statistical test, the result of the test is highly influenced by sample size.
Adverse Impact

- If an employer’s sample sizes are sufficiently large, even tiny differences in hiring rates will be statistically significant yielding a conclusion of adverse impact.
- 76% of male applicants are hired versus 75% of female applicants
- Given a large enough sample size, this tiny difference in hiring rates will be judged to be adverse impact.

Adverse Impact Triggers the Requirements of the Uniform Guidelines

- When an employment test results in adverse impact, the Uniform Guidelines make it the responsibility of the employer to have test validity documentation.
- Such validity documentation is expensive and labor intensive.
  - Often requires service of consulting firms and expert witnesses.

Adverse Impact Triggers the Requirements of the Uniform Guidelines

- The implicit assumption in the Uniform Guidelines is that adverse impact is an indication of a flawed test.
- The alternative explanation is that the test is an accurate assessment of sub-group differences in job-related KSAs.
  - By sub-group, one typically is referring to race or sex groups.
Sub-Group Difference in Job-Related KSAs are the Rule not the Exception

- There is a large body of evidence concerning race and sex differences in job-related KSAs.
- With the possible exception of personality tests, most employment tests show mean racial differences, on average.

Sub-Group Difference in Job-Related KSAs are the Rule not the Exception

- Cognitively-loaded tests (e.g., job knowledge, reading, math) will have the most race-based adverse impact, on average.
- Employment tests assessing spatial ability, mechanical comprehension, or upper body strength will have adverse impact by sex, on average.

Sub-Group Difference in Job-Related KSAs are the Rule not the Exception

- Sub-group differences in mean levels of job-related KSAs are the elephant in the room of personnel selection.
- An elephant in the room is an idiom referring to a large problem that is obvious but is not discussed.
- For any reasonable passing point on most any employment test, one can expect adverse impact, on average.
Sub-Group Difference in Job-Related KSAs are the Rule not the Exception

• This issue is uncomfortable to talk about and I don’t want to spend a lot of time on it.
• But, I have two sets of data.

High School Graduation Rates by Race

• Students who graduate with a diploma on time.
• Asian/Pacific Islander 91%
• Whites 81%
• American Indian/Alaskan Natives 64%
• Hispanics 64%
• Blacks 62%


High School Graduation Rates by Race

• These are large differences.
• Those who do not get a diploma, on average, have weak job skills (e.g., reading, writing, math).
• The US military has done a lot of research on those without a high school diploma.
  - Those without a high-school diploma have a substantially higher rate of leaving military service before their term is up.
High School Graduation Rates by Race

• In my opinion, American employers will have substantial adverse impact problems, as long as there are substantial disparities in high school diploma rates by race.

• In my opinion, rather than a local or state government spending money to meet requirements of the Uniform Guidelines, it would be better to spend money on programs designed to:
  - Increase job-related KSAs learned in high school
  - Increase high school graduation rates.

• In my opinion, private employers, if relieved of expensive validity documentation requirements of the Uniform Guidelines, could use some of that money to support programs to reduce racial disparities in high school diploma rates.
Prose literacy by race

- Intermediate level: Adults are able to read and understand moderately dense, less commonplace prose texts as well as summarize, make simple inferences, determine cause and effect, and recognize author’s purpose.

- Percent of adults at the intermediate level (2003 data):
  - Whites: 51%
  - Asians/Pacific Islander: 42%
  - Blacks: 31%
  - Hispanics: 23%


In my opinion, most knowledge-worker jobs require adults to read and understand moderately dense, less commonplace prose texts as well as summarize, make simple inferences, determine cause and effect, and recognize author’s purpose.

One typically needs these skills to graduate from college.

In my opinion, until such time as US mean racial differences in prose literacy are eliminated, we will have adverse impact problems on employment tests.
The Diversity-Validity Dilemma

- It is even more unpleasant.
- Scholars have recently become more honest in their discussion of the diversity-validity dilemma (e.g., Ployhart & Holtz, 2008).
- This dilemma is that the tests that best predict job performance tend to be the tests with the most race-based adverse impact.

The Diversity-Validity Dilemma

- Thus, the most valid tests (cognitively-loaded tests such as general mental ability, job knowledge, math, reading, reasoning) have the most adverse impact.
- The least valid tests (e.g., personality) have the least adverse impact.
- In sum, diversity goals and merit selection goals are typically in conflict.

Uniform Guidelines Flaw #6:

6a) Tests with adverse impact are the rule and not the exception.
6b) Tests with adverse impact tend to have larger validity than tests with less adverse impact.
6c) The onerous and expensive validity documentation requirements of the Uniform Guidelines should not be invoked solely because of adverse impact.
Differential Validity and Differential Prediction

- In the 1970's, the hypothesis of situational validity and the evidence that employment tests often yielded adverse impact fueled a debate concerning differential validity and differential prediction.
- This research explored whether the validity or the accuracy of the prediction of job performance varied by sub-group.

Because the situational specificity hypothesis held that some unknown characteristics controlled the magnitude of validity, it was considered possible that some unknown characteristics caused tests to be valid for one racial group but not another.
- Differential validity hypothesis

Local validity studies were encouraged as the best method to determine if the validities were the same for all sub-groups.
- However, by the mid-1980's, it became clear that differential validity was extremely rare.
- Much of this research was summarized in an influential National Academy of Sciences report.
Differential Validity and Differential Prediction

- Research on differential validity evolved into research on differential prediction.
- Even if the validity correlations were the same across all sub-groups, the optimal regression lines to predict job performance might differ.

Differential Prediction

- Differential prediction might occur either through different slopes or different intercepts.

Intercept & Slope Differences
Differential Prediction

• Research in the late 1970s and early 1980s showed that different regression slopes by group occurred at no higher than chance levels.
• Differing intercepts are more common and reflect group mean differences in job performance.

Typically when one has intercept differences, the error in prediction favors minority group members.
• That is, when intercepts vary by race, and the prediction of job performance is based on a common regression line (data on both mixed together), the performance of minority group members is overestimated.

Thus, to the extent that differential prediction exists, it benefits minority group members.
• Recently, Aguinis and colleagues (2010), noted that VERY large sample sizes are required to have adequate power to detect differential prediction.
Differential Prediction

- If Aguinis is correct, zero to very few employers in the US have a sufficient number of employees to conduct differential prediction studies.

Differential Prediction by Slope

- I suggest that differential prediction by slope is rare because it implies that a test measures one thing for one group and something else for another group.
- Consider a math problem.

Differential Prediction by Slope

- John had ten marbles and gave five marbles to Mary. How many marbles does John have left?
- If members of all sub-groups can read, this question measures exactly the same thing for members of most groups: subtraction.
- No differential prediction would be found.
Differential Prediction by Slope

• To find differential prediction by slope, this math question would need to measure math for one group but not for another.
• One needs to go to great lengths to come up with an example of how this could happen.

Differential Prediction by Slope

• Differential prediction could be manipulated into occurring by translating the math item into Korean. For those with reading comprehension in Korean, the item assesses math. For those without reading comprehension in Korean, it does not assess math.

Differential Prediction by Slope

• This example, although convoluted, is the only way that I see that one could get differential prediction by slope.
• No one has ever offered me a compelling example of why differential prediction by slope would occur in employment testing.
Differential Validity and Differential Prediction Summary

- There is no evidence that differential validity exists.
- Differential prediction may exist when two sub-groups have different mean levels of job performance (i.e., intercept differences).
- When this type of differential prediction exists, bias in prediction typically favors minorities.

Differential Validity and Differential Prediction Summary

- There is no evidence that differential prediction by slope exists in employment testing.
- There is evidence that differential prediction, if it existed, would be difficult to find given the huge number of employees required to find it.

Differential Prediction and the Uniform Guidelines

- Differential validity and prediction are addressed in the Uniform Guidelines under the concept of fairness.
- Because the Uniform Guidelines have not been updated with respect to scientific research on differential validity and differential prediction, the Uniform Guidelines encourage such analyses.
Uniform Guidelines Flaw #7: The Uniform Guidelines offer recommendations on differential validity and differential prediction that are at odds with scientific knowledge and professional practice as codified in the Standards and Principles.

Resistance to Changing the Uniform Guidelines

- The likely, yet unpleasant, reality is that the mean differences on employment tests reflect mean differences in job-related KSAs.
- Although mean sub-group differences are well known to researchers and practitioners in employment testing, it is impolite to talk of these differences.
  - Elephant in the room

Resistance to Changing the Uniform Guidelines

- It would likely be politically damaging to a politician or a political appointee (e.g., head of EEOC) to acknowledge that adverse impact is the norm and not the exception.
- Likewise, it would be politically damaging to acknowledge that employment tests, even with adverse impact, are generally accurate measures of job-related KSAs.
Resistance to Changing the Uniform Guidelines

- I suggest that the best opportunity to abolish or extensively revise the Uniform Guidelines lies within the Barrack Obama administration.
- Any white politicians suggesting the abolishment or extensive revision of the Uniform Guidelines would likely be called racists by some contributors to the debate.

Life after the Uniform Guidelines

- My comments here are necessarily speculative.
- In the absence of the Uniform Guidelines, it would be tough to argue that adverse impact is discrimination under the Civil Rights Act.
- This would sharply reduce the amount of employment litigation in this country.

Life after the Uniform Guidelines

- What would remain in litigation is employment discrimination due to disparate treatment.
  - Refuse to hire a female applicant because she is female.
- Employment in some areas will decline.
  - Attorneys and consultants in adverse impact cases would need to find other employment specialties.
Life after the Uniform Guidelines

- Guidance on good practices in personnel selection would be based on science and professional guidelines such as the Standards and the Principles.
- Employers’ administrative costs for employment testing will be lower because of fewer employment litigation concerns.

Life after the Uniform Guidelines

- Employer administrative cost savings may be diverted to improving personnel selection or to other activities that increase the effectiveness of the organization.

Life after the Uniform Guidelines

- Industrial and Organizational Psychologists can spend less time in litigation defense work and more time in improving the quality of selection methods and in exploring ways of reducing mean sub-group differences.
Life after the Uniform Guidelines

- The Federal government can shift its emphasis from threatening employers concerning adverse impact to enhancing education to reduce or eliminate the sub-group differences in job-related KSAs.
- Finally, the number of staff at the EEOC and the OFCCP can be sharply reduced and the resulting cost savings can be used to reduce the budget deficit.

Conclusion

- I think 107 slides could be viewed as Torture by Power Point and I should stop.
- It is my hope that my comments here have been viewed as constructive efforts to encourage debate concerning the flaws of the Uniform Guidelines.