33rd International Congress on Assessment Center Methods
London: September 2006

The Maturation of Validity Generalization (VG) in Defending Ability Assessment

James C. Sharf, PhD
Sharf & Associates
Employment Risk Advisors, Inc.
703/684-4886
jim@jimsharf.com

OUTLINE

Civil Rights Act of 1964

Supreme Court 1971: Griggs v. Duke Power

Schmidt & Hunter 1977: "accepted doctrine" of situational specificity???

EEOC's 1978 Uniform Guidelines on Employee Selection Procedures:

- situational specifity & single group validity

Evolution of Validity Generalization In Professional Standards and Principles:

- American Psychological Association's Standards for Educational & Psychological Tests, and
- The Society for Industrial and Organizational Psychology's (Division 14 of APA) Principles for the Validation and Use of Personnel Selection Procedures

TEXT (+ **powerpoints**)

1964: Civil Rights Act of 1964
"It shall be an unlawful employment practice for an employer: 1) to fail or refuse to hire or to discharge any *individual*, or otherwise to discriminate against an *individual* with respect to his compensation, terms, conditions, or privileges of employment because of such *individual's* race, color, religion, sex or national origin…" (emphasis added).

1971 Supreme Court: Griggs v. Duke Power Creates Group Rights
""Good intent or absence of discriminatory intent does not redeem employment procedures or testing mechanisms that operate as 'built-in headwinds' for minority groups and are unrelated to measuring job capability…. Congress directed the trust of the Act to the consequences of employment practices, not simply the motivation."

1971 Supreme Court: Griggs v. Duke Power "Finds" Group Rights
"Nothing in the Act precludes the use of testing or measuring procedures; obviously they are useful. What Congress has forbidden is giving these devices and mechanisms controlling force unless they are demonstrably a reasonable measure of job performance."

1971 Supreme Court: Griggs v. Duke Power Creates Group Rights
"Congress has not commanded that the less qualified be preferred over the better qualified simply because of minority origins. Far from disparaging job qualifications as such, Congress has made such qualifications the controlling factor, so that race, religion, nationality, and sex become irrelevant. What Congress has commanded is that any test used must measure the person for the job and not the person in the abstract."

1973: Equal Employment Opportunity Commission's Successful Advocacy Noted
"Even agencies which do not have the power to issue *substantive* regulations may have the power to issue *interpretative* regulations which will, in appropriate circumstances, be given weight by the courts. Indeed in Griggs, the Supreme Court held of EEOC's (Equal Employment Opportunity Commission) interpretation in the 1970 Guidelines that: 'The administrative interpretation of the Act by the enforcing agency is entitled to great deference'."

1976: Department of Justice Concedes Employment Testing Unduly Burdened
"Under the present (1970) EEOC guidelines, few employers are able to show the validity of *any* of their selection procedures, and the risk of their being held unlawful is high."

1976: Department of Justice Concedes Employment Testing Unduly Burdened, ctd.
"Since not only tests, but all other procedures must be validated, the thrust of the present guidelines is to place almost all test users in a position of non-compliance; to give great discretion to enforcement personnel to determine who should be prosecuted; and to set aside objective selection procedures in favor of numerical hiring."

1978 EEOC's Uniform Guidelines on Employee Selection Procedures
Embraced professional consensus of early 1970s based on Edwin Ghiselli's findings (1955 & 1966) regarding situational specificity and single group validity (a.k.a. test fairness, differential validity)

1955: Edwin Ghiselli: The Measurement of Occupational Aptitude
"As one looks through reports on the validity of any given type of test applied to workers on some particular job one notices the outstanding characteristic of variability. One author will present a validity coefficient that is quite low; and another author one that is substantial in magnitude…"

1955: Ghiselli: The Measurement of Occupational Aptitude, ctd.
"Undoubtedly, much of the variation results in differences in the requirements of the same job in different organizations."

1966: Ghiselli: The Validity of Occupational Aptitude Tests
"A given test applied to workers on a given job is very likely to have greater validity in one organization than in another. Indeed, determinations of the validity of a test for a given job, in one and the same organization, may vary significantly from one time period to another."

1966: Ghiselli: The Validity of Occupational Aptitude Tests, ctd.
"It is not easy to explain this variation… Certainly much of the variation results from differences in the nature of and the requirements for nominally the same job in different organizations, and in the same organization from one time period to another."

1975: Division 14 of American Psychological Association
Principles for the Validation and Use of Personnel Selection Procedures (Principles)
(Bob Guion & Mary Tenopyr)
"First, one must consider the rate of change of job characteristics… If technology, product, management practice, applicant populations, or other conditions are especially likely to change in ways that might affect validity, then a single criterion-related study is probably not useful…"

1975: Division 14 Principles
"Where traditional criterion-related validation is not feasible, the psychologist should consider alternative research strategies. These may involve techniques not yet well understood or sufficiently studies (such as) 'synthetic' validation… or generalization of validity."

1975: Division 14 Principles: Data Analysis
"If the candidate sample can be divided into sub-groups where prior research of logical considerations (such as indications of gross differences in job duties) suggest different relationships between predictors and criteria, the **analyses should be done separately for the sub-groups …(such) as ethnic group, socio-economic status, age, sex**, cognitive styles, etc., although this approach has generally not been found to improve validity appreciably."

1975: Division 14 Principles: Validity Generalization
"Validity evidence obtained in one unit of a multiunit organization or in a consortium may be applied to other units where jobs and job settings are essentially similar. **Validity coefficients are obtained in specific situations. They apply only to those situations.** Careful job and situational analyses are needed to determine whether characteristics of the site of the original research and those of other sites are sufficiently similar to make the inference of generalizability reasonable."

1975: Division 14 Principles: Validity Generalization
"A pressing problem in employment psychology is that of determining how to generalize validities. Psychologists are strongly urged to engage in cooperative research ventures such as industry-wide validation studies, consortia of civil service jurisdictions, and the like."

1975: Division 14 Principles: Validity Generalization
Until such time as such cooperative research results in an understanding of the limits of generalization, there will be few principles to observe in this area. The principle that one may apply validity evidence to essentially similar job units is an interim principle; it is not intended to discourage continued research to determine whether such application exceeds the legitimate boundaries of validity generalization."

1975: Division 14 Principles SUMMARY
Supported the "situational specificity" of validity
"Validity coefficients are obtained in specific situations. They apply only to those situations"

1975: Division 14 Principles SUMMARY
Supported "differential validity" (i.e., valid for one, but other "subgroups"??)
"analyses should be done separately for the sub-groups" (ethnic groups, sex)

1975: Division 14 Principles SUMMARY
Scientifically cautious about using validity generalization:
"Until such time as such cooperative research results in an understanding of the limits of generalization, there will be few principles to observe in this area"

1977: Schmidt & Hunter: Validity Generalization (VG)
Development of A General Solution to the Problem of Validity Generalization
"Sampling error leads to random variations in study outcomes, and measurement error artificially lowers… validities."

1977: Schmidt & Hunter: VG
"If all or most of the study-to-study variability in observed validities was due to artifacts, then the traditional belief in situational specificity would seem to be erroneous, and the conclusion would be that validity findings generalized."

1978: Equal Employment Opportunity Commission (EEOC) + Dept. of Labor's Office of Federal Contract Compliance Programs OFCCP) + Dept. of Justice + Civil Service Commission (→Office of Personnel Management)
Uniform Guidelines on Employee Selection Procedures (Uniform Guidelines)

1978: EEOC's Uniform Guidelines: "Situational Specificity"
Sec 7(B)(2) "Criterion-related validity studies conducted by one test user, or described in test manuals and the professional literature, will be considered acceptable for use by another user when… the incumbents in the user's job and the incumbents in the job …on which the validity study was conducted perform substantially the same major work behaviors… AND…"

1978: EEOC's Uniform Guidelines: "Differential / Single Group Validity"
Sec 7(B)(3) "The studies include a study of test fairness for each race, sex and ethnic group…"

1978: EEOC's Uniform Guidelines: "Unfairness"
Sec 14(B)(8)(a) "*Unfairness defined*: When members on one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores."

1980: Division 14 (→SIOP) of APA (Bill Owens & Mary Tenopyr)
Principles for the Validation and Use of Personnel Selection Procedures (2nd ed).
"Only that which is generalizable beyond the specific, immediate situation will have much meaning or practical use except to that specific situation. … Many questions regarding generalizability are still open to debate, but they are a matter of concern regardless of the validation strategy used."

1980: Division 14 Principles (2nd ed.): Validity Generalization.
"Classic psychometric teaching has long held that validity is specific to the research study and that inability to generalize is one of the most serious shortcomings of selection psychology."

1980: Division 14 Principles (2nd ed.) Procedural Considerations.
"***Where Traditional Criterion-Related Validaiton Strategy is Not Feasible, the Researcher Should Consider Any Alternative Research Methodology Which Offers a Sound Rationale*.** **Examples include synthetic validation, cooperative research on an industry-wide basis, consortia of small users, or gathering data for validity generalization**. However, the researcher should be aware that most non-traditional approaches require considerable research and development effort."

1980: Division 14 Principles (2nd ed.): Validity Generalization.
"**Current research is showing that the differential effects of numerous variables may not be as great as heretofore assumed**. To these findings are being added theoretical formulations, buttressed by empirical data, which propose that much of the difference in observed outcomes of validation research is due to statistical artifacts. Continued evidence in this direction should enable further extensions of validity generalization."

1980: Division 14 Principles (2nd ed.): Validity Generalization.
"Cooperative validation efforts being carried on by a number of trade and industry associations will provide the data necessary for evaluation. Such cooperative efforts are to be applauded and encouraged."

1980: Division 14 Principles (2nd ed.): Glossary
"Validity Generalization: The transportability of validity evidence; the application of validity evidence obtained in one or more situations to other situations."

1980: Division 14 Principles (2nd ed.) SUMMARY
Cautious acceptance of validity generalization:
***"Where Traditional Criterion-Related Validaiton Strategy is Not Feasible, the Researcher Should Consider Any Alternative Research Methodology Which Offers a Sound Rationale*.**
Examples include synthetic validation … or gathering data for validity generalization."

1980: Division 14 Principles (2nd ed.) SUMMARY
Still investigating the "situational specificity" of validity evidence:
"Only that which is generalizable beyond the specific, immediate situation will have much meaning or practical use except to that specific situation. … Many questions regarding generalizability are still open to debate, but they are a matter of concern regardless of the validation strategy used."

1980: Division 14 Principles (2nd ed.) SUMMARY
Questioning presumption of "differential validity":
"Current research is showing that the differential effects of numerous variables may not be as great as heretofore assumed."

1981: Schmidt & Hunter: VG
"Cognitive test validities can be generalized with confidence across settings, organizations, and there is no factual basis for requiring a validity study in each situation."

1981: Schmidt & Hunter: VG
"(C)umulative (meta) analyses show that the most frequently used cognitive ability tests are valid for all jobs and job families. In conclusion, our evidence shows that the validity of cognitive tests is neither specific to situations nor specific to jobs."

1982: Committee on Ability Testing in the National Research Council of the National Academy of Sciences: fairness
"Predictions based on a single equation (either one for whites or for a combined group of whites and blacks) generally yield predictions that are quite similar to, or somewhat higher than, predictions from an equation based only on data from blacks. In other words, the results do not support the notion that the traditional use of test scores in a prediction yields predictions for blacks that systematically underestimate their performance."

1985 APA (+AERA & NCME)
Standards for Educational and Psychological Testing: Validity Generalization
"In the past, judgments about the generalization of validity were often based upon nonquantitative reviews of the literature. In more recent years, quantitative meta-analytic techniques have been used frequently to study validity generalization. Both approaches have been used to support inferences about the degree to which validities generalize."

1985: APA Standards
"**These standards are concerned with a field that is evolving. Therefore, there is a continuing need for monitoring and revising this document as knowledge develops. There are some areas in which new developments are particularly likely, such as…. validity generalization**…"

1985: APA Standards: Validity Generalization
"Two uses of the results of validity generalization may also be distinguished: a) to draw scientific conclusions and b) to use the results of validity generalization obtained from prior studies to support the use of a test in a new situation. The latter use raises questions about the degree to which validities are transportable to a specific new situation."

1985: APA Standards: Validity Generalization
"An important issue in educational and employment settings is the degree to which criterion-related evidence of validity that is obtained in one situation can be generalized (that is transported and used) to another situation without further study of validity in the new situations. If generalization is limited, then local criterion-related evidence of validity may be necessary in most situations in which a test is used. **If generalization is extensive, then situation-specific evidence of validity may not be required**."

1985: APA Standards: Validity Generalization
"In conducting studies of the generalizability of validity evidence, the prior studies that are included may vary according to several situational facets. Some of the major facets are a) differences in the way the predictor construct is measured, b) the type of job or curriculum

involved, c) the type of criterion measure, d) the type of test takers, and e) the time period in which the study was conducted. In any particular study of validity generalization, any number of these facets might vary, and a major objective of the study is to determine whether variation in these facets affects the generalizability of validity evidence."

1985: APA Standards: Validity Generalization
"The extent to which predictive or concurrent evidence of validity generalization can be used as criterion-related evidence in new situations is in large measure a function of accumulated research. Consequently, although evidence of generalization can often be used to support a claim of validity in a new situation, the extent to which this claim is justified is constrained by the available data."

1985: APA Standards: Standard 1.16
"When adequate local validation evidence is not available, criterion-related evidence of validity for a specified test use may be based on validity generalization from a set of prior studies, provided that the specified test-use situation can be considered to have been drawn from the same population of situations on which validity generalization was conducted. (*Primary)"*

1985: APA Standards: Comment
"Several methods of validity generalization and simultaneous estimation have proven useful. In all methods, the integrity of the inference depends on the degree of similarity between the local situation and the prior set of situations. Present and prior situations can be judged to be similar, for example, according factors such as the characteristics of the people and job functions involved. Relational measures (correlations, regressions, success rates, etc.) should be carefully selected to be appropriate for the inference to be made."

1985: APA Standards: Standard 1.20
"Investigations of criterion-related validity for tests used in selection decisions should include, where feasible, a study of the magnitude of predictive bias due to differential prediction for those groups for which previous research has established a substantial prior probability of differential prediction for the particular kind of test in question.*"*

1985: APA Standards: Standard 1.21
"When studies of differential prediction are conducted, the reports should include regression equations (or an appropriate equivalent) computed separately for each group, job, or treatment under consideration or an analysis in which group, job, or treatment variables are entered as moderators. (*Primary*)"

1985: APA Standards: Standard 3.10
"When previous research indicates the need for studies of item or test performance differences for a particular kind of test for members of age, ethnic, cultural, and gender groups in the population of test takers, such studies should be conducted as soon as is feasible. Such research should be designed to detect and eliminate aspects of test design, content, or format that might bias test scores for particular groups."

1985: APA Standards: Standard 4.3
"Norms that are presented should refer to clearly described groups. These groups should be the ones with whom users of the test will ordinarily wish to compare the people who are tested. Test

publishers should also encourage the development of local norms by test users when the published norms are insufficient for particular test users."

1985: APA Standards: Comment
"When tests are developed for uses other than local use, the user needs to know the applicability of the test to different groups. Differentiated norms or summary information about differences between gender, ethnic, grade, or age groups, for example, may be useful. Users also need to be made alert to situations in which norms are less appropriate for one group than for another."

1985: APA Standards: Chapter 10. Employment Testing
"Employers should not be precluded from using a test if it can be demonstrated that the test has generated a significant record of validity in similar job settings for highly similar people or that it is otherwise appropriate to generalize from other applications."

1985: APA Standards: Glossary
"Validity generalization: Applying validity evidence obtained in one or more situations to other similar situations on the basis of simultaneous estimation, meta-analysis, or synthetic validation arguments."

1985: APA Standards SUMMARY
Cautious acceptance of validity generalization
"These standards are concerned with a field that is evolving. Therefore, there is a continuing need for monitoring and revising this document as knowledge develops. There are some areas in which new developments are particularly likely, such as…. validity generalization."

1985: APA Standards SUMMARY
Scientific skepticism of "situational specificity"
"If generalization is extensive, then situation-specific evidence of validity may not be required."

1985: APA Standards SUMMARY
Scientific skepticism of "situational specificity"
"Employers should not be precluded from using a test if it can be demonstrated that the test has generated a significant record of validity in similar job settings for highly similar people or that it is otherwise appropriate to generalize from other applications."

1985: APA Standards SUMMARY
Continued interest in investigating "single group validity"
"When tests are developed for uses other than local use, the user needs to know the applicability of the test to different groups. Differentiated norms or summary information about differences between gender, ethnic, grade, or age groups, for example, may be useful. Users also need to be made alert to situations in which norms are less appropriate for one group than for another."

1987: Society for Industrial / Organizational Psychology (Bill Owens & Neal Schmitt)
Principles for the Validation and Use of Personnel Selection Procedures: (3rd ed.)
"The Principles also contain discussions of the generality vs. specificity issue in validation. The need to develop selection procedures with generality is emphasized, not only to acknowledge practical considerations, but also to further the search for understanding of selection measures."

1987: SIOP Principles (3rd ed.): Determination of Feasibility:
"Validities appear to be quite stable across both tasks and situations, but there are influences such as restriction of range in the predictor, the criterion, or both, that may distort an estimate obtained from a particular sample."

1987: SIOP Principles (3rd ed.) The Researcher May Consider Alternate Criterion-Related Research Methods That Offer a Sound Rationale.
"Examples include cooperative research on an industry-wide basis, consortia of small users, or gathering data for validity generalization. Such approaches generally require considerable effort in planning and in data analysis and considerable care in ascertaining the appropriateness of the data included from different sources."

1987: SIOP Principles (3rd ed.) Design & Conduct of Criterion-Related Validity Studies
"The Sample for a Validation Study Should be Chosen Carefully. Whether such characteristics as age, race or sex affect predictor-criterion relationships is an empirical question, and the researchers should, therefore, rely on the research literature in making professional judgments about their possible relevance. Because many characteristics studied to date appear to have little or no effect on predictor-criterions relationships, no variable should be assumed to moderate validity coefficients in the absence of explicit evidence for such an effect. **The research literature shows that validities across races (black vs. white) are usually comparable on cognitive selection tests**."

1987: SIOP Principles (3rd ed.) Data Analysis: The Results Obtained in Criterion-Related Validity Studies Should Be Interpreted Against the Background of the Relevant Research Literature.
"Cumulative research knowledge plays an important role in any science. Large bodies of research regarding relationships between test scores and job performance currently exist. A researcher may, after consulting the literature, conclude that the existing research base is sufficient to support the use of certain instruments without any additional criterion-related research."

1987: SIOP Principles (3rd ed.) Data Analysis: The Results Obtained in Criterion-Related Validity Studies Should be Interpreted Against the Background of the Relevant Research Literature, ctd.
"In fact, in some instances, particularly when the sample of research subjects is small or unrepresentative, an additional criterion-related study may lead to an erroneous conclusion about the validity of a selection instrument. In any event, the results of an individual validity study should be interpreted in light of the relevant research literature. A history of similar findings in the research literature lends additional credence to the results of individual studies so long as the history is truly representative of the experience of the field. On the other hand, dissimilar findings in a single study should be viewed with caution."

1987: SIOP Principles (3rd ed.) Differential Prediction

"There is little evidence to suggest that there is differential prediction for the sexes, and the literature indicates that differential prediction on the basis of cognitive tests is not supported for the major ethnic groups. There is no compelling research literature or theory to suggest that cognitive tests should be used differently for different groups."

1987: SIOP Principles (3rd ed.): Validity Generalization

"Classic psychometric teaching has held that validity is specific to the research study and that inability to generalize is one of the most serious shortcomings of selection psychology. Current research has shown that the differential effects of numerous variables are not so great as heretofore assumed; much of the difference in observed outcomes of validation research can be attributed to statistical artifacts. **It now seems well established from both validity generalization studies and cooperative validation efforts that validities generalize far more than one supposed.**"

1987: SIOP Principles (3rd ed.)

"Differential prediction, in the sense of efforts to find different selection procedures for predicting different aspects of performance in a job or for predicting performance in different jobs, has proven to be difficult. This, too, is evidence that validities are more generalizable than has usually been believed."

1987: SIOP Principles (3rd ed.)

"To the extent that validity generalization evidence is available, researchers may rely on it to support the use of selection instruments. Researchers and employers are encouraged to conduct cooperative studies when adequate data for validity generalization are not available."

1987: SIOP Principles (3rd ed.)

"The cumulative validity evidence may indicate generalizability of validity for a selection procedure only for particular kinds of jobs or job families. In such a case, reliance on validity generalization results for jobs in new settings or organizations should meet two conditions: 1) The selection procedure to be used is a measure of the trait, ability, or construct studied, or is a representative example of the type of selection procedure included in the validity generalization study. … 2) The job in the new setting is similar to the job, or a member of the same job family, included in the validity generalization study… If sufficient information is available on the new job families to accurately assign it to a relevant occupational category, classification can be made without a formal job analysis."

1987: SIOP Principles (3rd ed.) SUMMARY

Endorsement of validity generalization

"It now seems well established from both validity generalization studies and cooperative validation efforts that validities generalize far more than one supposed"

1987: SIOP Principles (3rd ed.) SUMMARY

Endorsement of validity generalization

"To the extent that validity generalization evidence is available, researchers may rely on it to support the use of selection instruments."

1987: SIOP Principles (3rd ed.) SUMMARY
Rejection of "single group validity" (i.e., "differential prediction"
"The research literature shows that validities across races (black vs. white) are usually comparable on cognitive selection tests"

1987: SIOP Principles (3rd ed.) SUMMARY
Rejection of "single group validity" (i.e., "differential prediction"
"There is little evidence to suggest that there is differential prediction for the sexes, and the literature indicates that differential prediction on the basis of cognitive tests is not supported for the major ethnic groups. There is no compelling research literature or theory to suggest that cognitive tests should be used differently for different groups"

1987: SIOP Principles (3rd ed.) SUMMARY
Scientific skepticism of "situational specificity"
"Validities appear to be quite stable across both tasks and situations…"

1989: National Research Council of the National Academy of Sciences: VG
"We accept the general thesis of validity generalization, that the results of validity studies can be generalized to many jobs not actually studied, but we urge a cautious approach of generalizing validities only to appropriately similar jobs."

1989: National Research Council of the National Academy of Sciences: Fairness
"The use of a total-group regression equation generally would not give predictions that were biased against black applicants. If the total-group equation does give systematically different predictions than would be provided by the equation based on black employees only, it is somewhat more likely to overpredict than to underpredict. These results are generally consistent with results that have been reported for other tests."

1991: Civil Rights Act of 1991: Adverse Impact Definition Codified
"An unlawful employment practice … a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity …" (emphasis added).

1991: Civil Rights Act of 1991: "Race Norming" Prohibited
"(I)t shall be an unlawful employment practice for a respondent, in connection with the selection or referral of applicants or candidates for employment or promotion, to adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment related tests on the basis of race, color, religion, sex, or national origin."

1999: AERA (+APA & NCME) (Eva Baker & Paul Sackett)
Standards for Educational and Psychological Testing.
"For many testing applications, there is a considerable cumulative body of research that speaks to some, if not all, of the inferences … A meta-analytic integration of this research can form an integral part of the strategy for linking test information to the construct domain of interest. The value of collecting local validation data varies with the magnitude, relevance, and consistency of research findings using similar predictor measures and similar criterion construct domains for similar jobs. In some cases, a small and inconsistent cumulative research record may lead to a

validation strategy that relies heavily on local data; in others, a large, consistent research base may make investing resources in additional local data collection unnecessary."

1999: AERA Standards: Standard 14.2
"When a test is used to predict a criterion, the decision to conduct local empirical studies of predictor-criterion relationships and interpretation of the results of local studies of predictor-criterion relationships should be grounded in knowledge of relevant research."

1999: AERA Standards: Comment
"**The cumulative literature on the relationship between a particular type of predictor and type of criterion may be sufficiently large and consistent to support the predictor-criterion relationship without additional research**. In some settings, the cumulative research literature may be so substantial and so consistent that a dissimilar finding in a local study should be viewed with caution unless the local study is exceptionally sound. Local studies are of greatest value in settings where the cumulative research literature is sparce (e.g., due to the novelty of the predictor and/or criterion used), where the cumulative record is inconsistent, or where the cumulative literature does not include studies similar to the local setting (e.g., a test with a large cumulative literature dealing exclusively with production jobs and a local setting dealing with managerial jobs."

1999: AERA Standards SUMMARY
Endorsement of validity generalization
"The cumulative literature on the relationship between a particular type of predictor and type of criterion may be sufficiently large and consistent to support the predictor-criterion relationship without additional research."

2003: SIOP (Dick Jeanneret)
Principles for the Validation and Use of Personnel Selection Procedures (4th ed.).
Generalizing Validity Evidence
"**At times, sufficient accumulated validity evidence is available for a selection procedure to justify its use in a new situation without conducting a local validation research study.** In these instances, use of the selection procedure may be based on demonstration of the generalized validity inferences from that selection procedure, coupled with a compelling argument for its applicability to the current situation. Although neither mutually exclusive nor exhaustive, several strategies for generalizing validity evidence have been delineated: (1) transportability, (2) synthetic validity/job component validity, and (3) meta-analytic validity generalization."

2003: SIOP Principles (4th ed.): Transportability
"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 applicant group."

2003: SIOP Principles (4th ed.): Synthetic Validity/Job Component Validity

"A second approach to establishing generalized evidence of the validity of inferences based on scores from a selection procedure is referred to as synthetic validity or job component validity. (While some researchers distinguish these terms, others do not, and in either case several variations on each exist.) A defining feature of synthetic validity/job component validity is the justification of the use of a selection procedure based upon the demonstrated validity of inferences from scores on the selection procedure with respect to one or more domains of work (job components). Thus, establishing synthetic validity/job component validity requires documentation of the relationship between the selection procedure and one or more specific domains of work (job components) within a single job or across different jobs. If the relationship between the selection procedure and the job component(s) is established, then the validity of the selection procedure for that job component may be generalizable to other situations in which the job components are comparable."

2003: SIOP Principles (4th ed.): Synthetic Validity/Job Component Validity, ctd.

"The validity of a selection procedure may be established with respect to different domains (components) of work, then "synthesized" (combined) for use based on the domains (or components) of work relevant for a given job or job family. In some instances, this may involve conducting a research study designed to demonstrate evidence for the generalized validity of inferences from scores on a set of selection procedures, and then using various subsets of these procedures for selection into both jobs or job families in the original study as well as into other jobs or job families. In other cases, it may involve generalizing the validity of inferences based on scores on selection procedures examined in one or more research studies conducted elsewhere to the new situation. In both cases, detailed analysis of the work is required for use of this strategy of generalizing validity evidence."

2003: SIOP Principles (4th ed.): Meta-Analysis

"Meta-analysis … can be used to determine the degree to which predictor-criterion relationships are … generalizable to other situations… Meta-analysis requires the accumulation of findings from a number of validity studies to determine the best estimates of the predictor-criterion relationship for the kinds of work domains and settings included in the studies…" "The question to be answered using a meta-analytic strategy is whether the valid inferences about work behavior or job performance can be drawn from predictor scores across given jobs or job families in different settings. (Note that the focus here is on using meta-analysis to examine predictor-criterion relationships...)."

2003: SIOP Principles (4th ed.): Meta-Analysis, ctd.

"**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 … 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 non-cognitive constructs in employment settings**."

2003: SIOP Principles (4th ed.): Meta-Analysis, ctd.

"Reports that contribute to the meta-analytic research results should be clearly identified and available. Researchers should consult the relevant literature to ensure that the meta-analytic

strategies used are sound and have been properly applied, that the appropriate procedures for estimating predictor-criterion relationships on the basis of cumulative evidence have been followed, that the conditions for the application of meta-analytic results have been met, and that the application of meta-analytic conclusions is appropriate for the work and settings studied. The rules by which the researchers categorized the work and jobs studied, the selection procedures used, the definitions of what the selection procedure is measuring, the job performance criteria used, and other study characteristics that were hypothesized to impact the study results should be fully reported. The quality of the individual research studies and their impact, if any, on the meta-analytic conclusions and their use also should be informed by good professional judgment ..."

2003: SIOP Principles (4th ed.): Meta-Analysis, ctd.
"**Generalizing validity evidence from meta-analytic results is often more useful than a single study**. However, 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), a local individual study may be more accurate than the average predictor-criterion relationship reported in a meta-analytic study. A competently conducted study, with a large sample using the same test, for the same kind of work activities, may be more accurate, informative, and useful than a cumulation of small validation studies that are not representative of the setting to which one wants to generalize validity."

2003: SIOP Principles (4th ed.): Predictive bias.
"Predictive bias has been examined extensively in the cognitive ability domain. For White-African American and White-Hispanic comparisons, slope differences are rarely found; while intercept differences are not uncommon, they typically take the form of overprediction of minority group performance."

2003: SIOP Principles (4th ed.): SUMMARY
Rejection of "Unfairness"
"Predictive bias has been examined extensively in the cognitive ability domain. For White-African American and White-Hispanic comparisons, slope differences are rarely found; while intercept differences are not uncommon, they typically take the form of overprediction of minority group performance."

2003: SIOP Principles (4th ed.): SUMMARY
Endorsement of validity generalization
"At times, sufficient accumulated validity evidence is available for a selection procedure to justify its use in a new situation without conducting a local validation research study."

2003: SIOP Principles (4th ed.): SUMMARY
Endorsement of validity generalization
"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 … These findings are particularly well-established for cognitive ability tests."

2003: SIOP Principles (4th ed.): SUMMARY
Endorsement of validity generalization
"Generalizing validity evidence from meta-analytic results is often more useful than a single study."

CHRONOLOGY: Standards & Principles

1954: APA
Technical Recommendations for Psychological Tests and Diagnostic Techniques

1955: NEA (AERA & NCME)
Technical Recommendations for Achievement Tests

1966: APA (AERA & NCME)
Standards for Educational and Psychological Tests and Manuals

1974: APA (AERA & NCME)
Standards for Educational & Psychological Tests

1975: Division 14 of APA (Bob Guion & Mary Tenopyr)
Principles for the Validation and Use of Personnel Selection Procedures.

1980: Division 14 of APA (Bill Owens & Mary Tenopyr)
Principles for the Validation and Use of Personnel Selection Procedures (2nd ed).

1985 APA (AERA & NCME)
Standards for Educational and Psychological Testing.

1987: SIOP (Bill Owens & Neal Schmitt)
Principles for the Validation and Use of Personnel Selection Procedures: (3rd ed.)

1999: AERA (APA & NCME) (Eva Baker & Paul Sackett)
Standards for Educational and Psychological Testing.

2003: SIOP (Dick Jeanneret)
Principles for the Validation and Use of Personnel Selection Procedures (4th ed.).

BIBLIOGRAPHY

Ghiselli, E. (1955). The Measurement of Occupational Aptitude. Berkeley, CA: University of California Press.

Ghiselli, E. (1966). The Validity of Occupational Aptitude Tests. New York, NY: John Wiley & Sons.

Guion, R.M. (1991). Personnel assessment, selection and placement. In Dunnette, M., & Hough, L. (Eds.). Handbook of Industrial & Organizational Psychology. Palo Alto, CA: Consulting Psychology Press.

Guion, R.M. (1998). Assessment, Measurement and Prediction for Personnel Decisions. Mahwah, N.J.: Lawrence Earlbaum.

Hartigan, J.A., & Wigdor, A.K. (Eds.) (1989). Fairness in Employment Testing: Validity Generalization, Minority Issues and the General Aptitude Test Battery. Washington, D.C.: National Academy Press.

Hough, L. M., Oswald, F. L., & Ployhart, R. R. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment, 9,* 1-43.

Hunter, J.E., & Schmidt, F.L. (1966). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*. 1, 199-223.

Murphy, K. (Ed.) (2003). Validity Generalization: A Critical Review. Mahwah, N.J.: Lawrence Earlbaum.

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.

Slagado, J., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rollland, J. (2003). A meta-analytic study of general mental ability for different occupations in the European community. *Journal of Applied Psychology*, 88, 1068-1081.

Schmidt, F.L. (1992). What do data really mean? Research findings, meta-analysis and cumulative knowledge in psychology. *American Psychologist*, 47, 1173-1181.

Schmidt, F.L., & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 643-661.

Schmidt, F.L., & Hunter, J.E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128-1137.

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.

Schmidt, F.L., Hunter, J.E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. *Journal of Applied Psychology*, 66, 166-185.

Schmidt, F.L., Hunter, J.E., Pearlman, K., & Hirsch, H. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology*, 38, 697-798.