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: (a) transportability, (b) synthetic validity/job component validity, and (c) meta-analytic validity generalization.

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 candidate group.

Synthetic Validity/Job Component Validity

A second approach to establishing generalized 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.

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.

Meta-Analysis

Meta-analysis is a third procedure and strategy that can be used to determine the degree to which predictor-criterion relationships are specific to the situations in which the validity data have been gathered or are generalizable to other situations, as well as to determine the sources of cross-situation variability (Aguinis & Pierce, 1998). 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.

While transportability and synthetic validity/job component validity efforts may be based on an original study or studies that establish the validity of inferences based on scores from the selection procedure through a content-based and/or a criterion-related strategy**,** meta-analysis is a strategy that only can be applied in cases in which the original studies relied upon criterion-related evidence of validity**.** 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. Meta-analysis also can be used to examine other issues, such as convergence among instruments intended to measure the same construct.)

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

Professional judgment in interpreting and applying the results of meta-analytic research is important. Researchers should consider the meta-ana-

lytic methods used and their underlying assumptions, the tenability of the assumptions, and artifacts that may influence the results (Bobko & Stone-Romero, 1998; Raju, Anselmi, Goodman, & Thomas, 1998; Raju et al., 1991; Raju, Pappas, & Williams, 1989). In evaluating meta-analytic evidence, the researcher should be concerned with potential moderators to the extent that such moderators would affect conclusions about the presence and generalizability of validity. In such cases, researchers should consider both statistical power to detect such moderators and/or the precision of estimation with respect to such moderators. In addition, the researcher should consider the probabilities of both Type I and Type II decision errors (Oswald & Johnson, 1998; Sackett, Harris, & Orr, 1986).

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 (Guion, 1998; Law, Schmidt, & Hunter, 1994a, 1994b).

Note that sole reliance upon available cumulative evidence may not be sufficient to meet specific employer operational needs such as for the placement of employees or for the optimal combination of procedures. Consequently, additional studies and data may be required to meet these specific needs. If such studies are not feasible in an organization, researchers and employers may engage in cooperative studies.

Meta-analytic methods for demonstrating generalized validity are still evolving. Researchers should be aware of continuing research and critiques that may provide further refinement of the techniques as well as a broader range of predictor-criterion relationships to which meta-analysis has been applied.

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.

Reliance on meta-analytic results is more straightforward when they are organized around a construct or set of constructs. When different predictors (as well as different criteria) intended to measure the same construct are combined in a meta-analysis, findings are meaningful to the extent that there is evidence that they do indeed reflect the same construct (e.g., convergent validity evidence). If, for example, meta-analytic evidence relies on data from five highly correlated, published measures of a predictor construct, these findings cannot be assumed to generalize to other measures using the same construct label without evidence that those other measures indeed reflect the same construct.

When studies are cumulated on the basis of common methods (e.g., interviews, biodata) instead of constructs, a different set of interpretational difficulties arise. Generalization is straightforward when, for example, an empirical biodata scale has been developed for a specific occupation, multiple validity studies have been conducted using that scale in that occupation, and the intent is to generalize to another setting that employs individuals in that same occupation. However, researchers may have difficulty when they attempt to generalize about a method in general, rather than about a specific application of the method. Because methods such as the interview can be designed to assess widely varying constructs (from job knowledge to integrity), generalizing from cumulative findings is only possible if the features of the method that result in positive method-criterion relationships are clearly understood, if the content of the procedures and meaning of the scores are relevant for the intended purpose, and if generalization is limited to other applications of the method that include those features. Consider the cumulation of validity findings from various interview methods, where in the population of settings in which interviews are used, the interview development process, content, and scoring vary (e.g., some knowledge-focused and some value-focused; some structured and some unstructured). Now consider a setting in which these features have not been coded, and thus it is unclear whether these features vary in the sample of studies available for meta-analysis. Generalizing from a meta-analysis of such data to a new similarly unspecified interview, to a different interview method, or to a different or new situation, is not warranted. For example, it may be the case that all studies in the database involve knowledge-focused interviews, and consistency in validity across knowledge-focused interviews offers no grounds for inferring that validity will generalize to value-focused interviews. In contrast, a cumulative database on interviews where content, structure, and scoring are coded could support generalization to an interview meeting the same specifications.