Practical Assessment, Research, and Evaluation

1990

The Case for Validity Generalization

Fred Rafilson

Follow this and additional works at: https://scholarworks.umass.edu/pare

Practical Assessment, Research & Evaluation

The Case for Validity Generalization

Rafilson, Fred
Wonderlick, Inc.

An important issue in educational and employment settings is the degree to which evidence of validity obtained in one situation can be generalized to another situation without further study of validity in the new situation. The issue of Validity Generalization is discussed in this article. Theory, procedures, and applications are addressed.

The extent to which predictive or concurrent evidence of validity can be used as criterion-related evidence in new situations is, in large measure, a function of accumulated research. In the past, judgments about the generalization or transportability of validity were often based on nonquantitative reviews of the literature. Today, quantitative techniques have been more frequently employed to study the generalization of validity (Schmidt, Hunter, Pearlman, & Hirsh, 1985). Both approaches have been used to support inferences about the degree to which the validity of a given predictor variable can generalize from one situation or setting to another similar set of circumstances.

If validity generalization evidence is limited, then local criterion-related evidence of validity may be necessary to justify the use of a test. If, on the other hand, validity generalization evidence is extensive, then situation-specific evidence of validity may not be required.

THEORY

A major limitation to local validation studies is that they can readily suffer from unseen local methodological problems. By comparing validation and fairness findings across multiple studies, however, it is possible to determine if the criterion-related validity of a test is relatively stable or if the test is valid only in certain situations. Drawing on meta-analysis techniques, this comparative procedure is called validity generalization in the personnel selection and psychometric literature.

Several types of measures lend themselves particularly well to validity generalization. Meta-analyses of the plethora of validity studies conducted on general cognitive ability (g) have repeatedly shown that the validity of g for predicting success in a given job differs little from one setting to another (Schmidt & Hunter, 1981). Thus, there is significant evidence that the validation results for general cognitive ability measures are generalizable across settings. It is not necessary, therefore, to conduct a validity study for a given job at every business location in America. The validity of 'general cognitive ability' for predicting clerical performance in one setting, for example, can be inferred from the validity found in the hundreds of previous studies.

Another limitation of specific local validation studies is the accuracy of the generated statistics (Schmidt, Hunter & Urry, 1976). Accurate statistics require large sample sizes. The criterion related validity of a test in a local validation study is usually inferred only if the findings reach a certain level of magnitude called 'statistical significance'. The smaller the sample of subjects, the higher the observed validity coefficient would need to be in order to infer an acceptable level of validity.

You would not expect, for example, to draw accurate predictions of a national election by polling a sample of only 15 voters. Most polls interview 1,000 voters or more. The same is true of the statistics produced by a local validation study; there is huge sampling error in individual validation studies conducted with small samples. Unless there are hundreds of subjects at a particular location, the data cannot be used to draw accurate conclusions in isolation. Rather, the data from small local samples can only be used cumulatively by combining them with the results from other local studies as is done in a validity generalization study.

PROCEDURE

In conducting validity generalization studies, data used from local studies may vary according to several situational facets. These may include:

- differences in the way the predictor construct are measured;
- the type of job or curriculum involved;

- the type of criterion measure;
- the type of test takers; and
- the time period in which the study was conducted.

In any particular validity generalization study, any number of these facets may vary. A major objective of the study is to determine whether variation in these facets affects the generalizability of validity evidence.

A common procedure for conducting a meta-analysis to determine the degree to which validity findings can be generalized is to

a) estimate the population validity by computing the mean of the observed sample validities,

b) correct the observed validities by removing the effects of statistical artifacts (Four readily quantifiable artifacts which can be controlled statistically are: sampling error, criterion unreliability, range restriction, and predictor unreliability),

c) find the variance of the corrected observed validities (the residual variance of the observed correlations after removing the statistical artifacts).

If the variance of the corrected observed validity is nearly zero, then validity generalizes and can be transported to other situations or locations.

MODELS

At present there are three different models for assessing Validity Generalization:

- the correlation model,
- the covariance model, and
- the regression slope model.

A recent empirical Monte Carlo study (Raju, Williams, & Pappas, 1989), conducted with an extremely large database (N=84,808), showed that all three models perform similarly. The regression slope model, however, may be more robust in some situations when the metrics for the predictor and the criterion can be considered comparable across studies.

APPLICATIONS

There are two main uses of validity generalization studies. First, the results of generalization studies can serve to draw scientific conclusions about the relationships between variables. A good example of this application is the conclusion drawn by Hunter and Schmidt (1981) that "the most frequently used cognitive ability tests are valid for all jobs and all job families...that the validity of the cognitive tests studied is neit her specific to situations or specific to jobs." In turn, these findings can improve our understanding of the true test/criterion relationships, allowing for a more useful application of predictor scores.

Second, the evidence of criterion related validity obtained from prior studies can be used to support the use of a test in a new situation. This application of validity generalization theory has enormous potential for educators and employers who lack sufficient sample sizes or resources in a given organization, yet would like to implement a proven valid testing program. This 'transference' of a test from one situation in which the test has been proven valid to another similar situation or location is often referred to as the 'transportability' of validity from one situation to another.

REFERENCES

Raju, N.S., Williams, C.P., & Pappas, S., (1989), An empirical monte carlo test of the accuracy of the correlation, covariance, and regression slope models for assessing validity generalization. *Journal of Applied Psychology*, 74, 901911.

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., Pearlman, K., & Hirsh, H.R. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology*, 38, 697-798.

Schmidt, F.L., Hunter, J.E., & Urry, V.W. (1976), Statistical power in criterion-related validity studies. *Journal of Applied Psychology*, 61, 473-485.

———