

TECHNICAL BRIEF

Hogan Research Methodology

Chapter 4: Criterion-Related Validity Evidence

- ❖ Concurrent Criterion Related Validity Study
- ❖ Predictive Criterion Related Validity Study

This technical brief is an excerpt of the Criterion-Related Validity chapter from the Hogan Research Methodology (HRM) addendum for all technical documentation.

A4. CRITERION-RELATED VALIDITY EVIDENCE

Aguinis, Henle, and Ostroff (2001) described criterion-related validity in terms of the relationship between the predictor (e.g., HPI Scales) and some criterion measure (e.g., job performance), with the goal of answering the basic question: how accurate are test scores in predicting criterion performance? The *Uniform Guidelines* state “evidence of the validity of a test or other selection procedure by a criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance” (29 C.F.R. § § 1607.5 (B)).

Although there are many organizational and logistical constraints that limit the usefulness of criterion-related validity studies (McPhail, 2007), the *Uniform Guidelines* and *Principles* suggest considering this approach when a) there is an adequate, representative sample of job incumbents willing to participate, and b) development of reliable, unbiased measures of job performance is possible. The *Principles* also recommends using a relevant criterion measure, one that “reflects the relative standing of employees with respect to important work behavior(s) or outcome measures(s)” (p. 14).

Additional factors should be taken into account and used as a guide when determining whether a criterion-related validity study is appropriate to use in any given selection situation. First, practitioners should consider the design when planning the study. A predictive design predicts scores on a criterion measured at some future occurrence. For example, *job applicants* complete the assessment before being hired and provide measures of performance after being on the job for some time. Concurrent designs are more practical because they do not require a time delay; instead, the organization collects job performance information at the same time *job incumbents* complete the assessment battery. Only one empirical study has examined the effects of these two strategies on criterion-related validity using personality measures. Van Iddekinge and Ployhart’s (2008) review of criterion study design revealed that predictive designs produce slightly lower validity estimates than concurrent designs. Yet regardless of the strategy employed, the predictive value of the assessment is established by correlating assessment scores and job performance data, and other factors beyond study design may still influence this validity coefficient.

For example, the *Principles* note that this observed validity coefficient “may underestimate the predictor-criterion relationship due to the effects of range restriction and unreliability in the predictors and/or criterion.” As a result, adjustments are available to account for these artificial reductions in variance. For instance, researchers often correct for criterion unreliability to estimate operational validity (Van Iddekinge & Ployhart, 2008). Note that Hogan corrects for measurement error, criterion unreliability, and range restriction where appropriate and reports both the observed and corrected validity coefficients in our technical documentation. Note that we do not correct correlation coefficients for predictor unreliability to estimate validity at the construct level

Another decision researchers face is whether to use a single criterion or multiple criteria during the data collection phase of the criterion study. The literature recommends that researchers “develop criterion measures that are conceptually aligned with the latent criterion constructs and that maximize the potential use of multiple criteria for predictor validation”

(Van Iddekinge & Ployhart, 2008, p. 906). Furthermore, J. Hogan and Holland (2003) provide strong support for using specific criteria to estimate the validity of specific predictors in operational use. Although support for using narrow criteria is growing, collecting overall performance composites still provide the best approach to estimating validity of global predictors (Guion, 1961) and prediction improves when criterion ratings cover the full spectrum of effective performance (Oh & Berry, 2009). Therefore, the use of global criteria in the design of performance rating forms is still appropriate; however, specific dimensions should also be used when circumstances allow for them.

Adhering to these guidelines, Hogan conducts criterion-related validity studies when an organization can (a) identify enough incumbents (e.g., generally greater than 100) to take the assessments and (b) identify supervisors to evaluate each incumbent using a performance rating form developed by Hogan. Many organizations choose not to conduct a local validation when they either do not have enough incumbents in a role or are concerned about the time commitment involved for their incumbents and their supervisors. In such cases, Hogan relies solely on VG evidence.

A4.1. Concurrent Criterion Related Validity Study

Hogan conducts concurrent criterion-related validity studies using a three-step process: (1) collecting Hogan assessment data, (2) collecting job performance data (i.e., supervisor and objective performance ratings), and (3) conducting analyses examining the relationships between the assessment and performance data.

In step one, the organization and Hogan work together to identify incumbents to participate in the research. Ideally, this sample should include at least 100 incumbents where we obtain matched predictor and criterion data. Additionally, this sample should represent the incumbent population in terms of performance, demographics, locations, work environments, and other contextual factors that may influence job performance. Last, incumbents should have enough tenure in the position to allow for reliable performance ratings by a supervisor. Identified incumbents then complete the relevant Hogan assessments.

In step two, Hogan uses information gathered from the job analysis and stakeholder conversations to develop a Performance Rating Form (PRF). This PRF usually includes items to assess overall performance, critical competencies, and other relevant behaviors. Incumbents' supervisors often participate in rater training (i.e., to reduce bias rating errors) prior to completing the PRF. This process may also include collecting objective performance metrics (e.g., sales, turnover, business unit performance) from the organization to help supplement the data available for step three.

In the final step, Hogan analyzes the incumbent assessment and performance data to identify the assessments and scales that are the strongest predictors of performance in the job. Hogan uses this information to create a multi-scale profile of successful performance on the job. The client may then implement this profile to aid in their candidate selection process.

A4.2. Predictive Criterion Related Validity Study

Predictive criterion related validity studies closely resemble concurrent criterion-related validity studies with one notable exception: the sample. The sample for predictive criterion related validity studies consists of job applicants, as opposed to job incumbents (Anastasi & Urbina, 1997). As such, the process for conducting a predictive study varies slightly from a concurrent study.

The first step in a predictive study involves administering the measure to job applicants. Candidates are then selected for the role without considering results from the administered measure. At a later date, Hogan collects performance data based on job analysis information for the hired applicants. Hogan analyzes the relationship between the administered scales and job performance to identify relevant scales for predicting job performance. Given practical and operational considerations, companies rarely have the opportunity to conduct this type of study and often rely on Hogan to use a concurrent design (as discussed in section A4.1).

REFERENCES

- Aguinis, H., Henle, C. A., & Ostroff, C. (2001). Measurement in work and organizational psychology. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 27–50). London: Sage
- Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.
- Guion, R. M. (1961). Criterion measurement and personnel judgments. *Personnel Psychology*, *14*, 141-149.
- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, *88*, 100-112.
- McPhail, S. M. (2007). *Alternative validation strategies*. San Francisco, CA: Jossey-Bass.
- Oh, I., & Berry, C. M. (2009). The Five-Factor model of personality and managerial performance: Validity gains through the use of 360 degree performance ratings. *Journal of Applied Psychology*, *94*, 1498-1513.
- Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology*, *61*, 871-925.