Guidelines for Criterion-related Validation Studies

Purpose

The purpose of a criterion-related validation study is to provide validity evidence to support the effectiveness of a selection tool in the form of a statistical correlation between the test (predictor) and job performance (criterion).

Acceptable Methods

There are two primary methods for conducting criterion-related validation studies. A predictive methodology requires that the predictor test be administered to a group of applicants for data collection purposes; however, the applicants are selected based on the use of a separate selection tool. At a later time period, criterion data can be collected for each applicant who was selected, and a correlation can be determined between predictor and criterion. This methodology is both time-consuming and impractical in most cases.

A more useful methodology is the concurrent validation model. This model requires that the predictor test be administered to an incumbent sample and that criterion data be collected from this same sample. As a general rule, I/O Solutions prefers the concurrent methodology based on its speed and efficiency.

Criterion Measurement

The purpose of a criterion variable is to create a metric for evaluating job performance. For I/O Solutions' purposes, there are generally two criterion variables that will be considered acceptable for a criterion-related validation study. The preferred variable is an assessment of global job performance. This can take the form of a standardized job performance appraisal rating or a forced rank-ordering of global incumbent performance. A secondary variable that can be used is a global evaluation of training or probationary performance. Academic performance in a training academy or probationary performance evaluations completed by direct supervisors are sufficient for estimating the expected performance of incumbent personnel (this is the case because success in the academy or during a probationary training period are prerequisites of success on the job).

Constraints (range restriction, performance data availability, union issues, range in performance)

In any attempt to measure the true relationship between a test and criterion, we recognize that a myriad of variables exist that will moderate, attenuate or otherwise impede the criterion-related validation process. Following is a brief description of these impediments and suggestions for managing these variables:

- Union buy-in: Due to the involvement of job performance data, many collective bargaining bodies reject the notion of participating in a validation process. I/O Solutions has found that speaking directly with union officials and describing the purpose of test validation and its contribution to selecting high-quality employees contributes to greater understanding and acceptance.

Copyright © 2010, IOS, Inc.

- Sampling: It is critical that a validation study be based on a representative and sufficiently large sample of incumbents. The sample cannot be composed of only high performers or only volunteers. The sample should be randomly selected and consist of a significant percentage of the overall workforce.
- Exam score reliability/accuracy: When using a concurrent methodology that involves field-testing, often incumbents do not take the test seriously and fail to perform at their true ability level. It is critical for supervisors/managers to explain the need for compliance and to motivate incumbents to take their time and perform at a high level. It is also necessary for the agency to provide incumbents the necessary time to complete this exercise so that the incumbent is neither rushed nor participating without compensation.
- Existence of performance data: Often there does not exist a performance evaluation tool or archival data that can be used in the validation study. In such a case, I/O Solutions can provide a standardized performance evaluation that can be completed by direct supervisors.
- Performance appraisal quality/range: It is typical for performance evaluations to result in the majority of employees receiving high ratings and for these rating to be inaccurate. In such a case, I/O Solutions will provide a standardized performance evaluation tool that can be completed by direct supervisors. I/O Solutions may also request that a group of supervisors convene to rank order or classify all employees based on performance. This methodology will be employed where it is necessary to force variance in the rating process.
- Intended use of performance data: It is a common concern about subordinates that performance data collected for a criterion-related validation study will be used for evaluative or punitive purposes. It is critical to the acceptance of the process that these data remain confidential and only be used for the purpose of the validation study. This point should be clearly communicated to incumbents.
- Predictor range restriction: A major fault of the criterion-related validation study is that it is limited to sampling individuals who were hired and therefore have performed successfully on the tests that were employed as selection devices. This creates a restriction in range among test scores. The only solution to this problem is to enact a statistical correction to account for such attenuation.

Sample Size

In order for a criterion-related validation study to be meaningful, it should be based on a sufficiently large sample. The selection of a proper sample size for a behavioral research study involves a decision regarding the statistical power that is desired for the inferences that will be drawn from that study. Specifically, a researcher must determine the level of confidence and the risk of error that is acceptable for the study. Sampling error is the error that is caused by drawing conclusions based on a sample rather than an entire population. The basic premise is that a conclusion is most accurate when it is based on all possible data points and that the conclusion becomes less accurate as less of the population is represented. Both sample size and sampling method will affect the quality of the conclusions that are drawn during research. Roscoe (1975) suggested some simple rules of thumb for selecting appropriate sample sizes based on an

Copyright © 2010, IOS, Inc.

analysis of acceptable confidence levels in behavioral research studies. The general recommendation is that sample sizes be at least 30 and need not be larger than 500 (at 500, sample error will not exceed 10 percent of the standard deviation about 98 percent of the time). Further, within this range of 30 to 500, it is appropriate to sample 10 percent of a parent population (Alreck & Settle, 1995). Therefore, if an agency contains a population of 500 people, it would be acceptable to conduct a criterion-related validation study on a sample of 50. A split-half analysis of consistency will serve as a prudent check of the adequacy of the sample size by ensuring that two randomly selected halves of the data produce similar results. If a split-half analysis is to be performed, you would need to double the size of your target sample.

Attenuation Correction

Attenuation is the weakening of a relationship between variables due to measurement error. In other words, our inability to perfectly sample a population and control for moderating variables results in an imperfect picture of the relationship or correlation between variables (namely a test and a measure of job performance). I/O Solutions recognizes two corrections for validity coefficient attenuation that are acceptable to produce a more accurate understanding of the true strength of the predictor-criterion relationship. The first correction, where necessary, is for unreliability in the criterion. A conservative estimate of criterion reliability, or a calculated estimate when possible, will be used to facilitate this correction. The second acceptable correction is for range restriction in the predictor. I/O Solutions will use our large database of normative data to define normal candidate/applicant mean and standard deviations. Predictor statistics that fail to fall in line with normative statistics will be corrected for range restriction.

Rules of Thumb for Judging Coefficient Magnitude

The following table provides guidelines from the U.S. Department of Labor's *Testing and Assessment, An Employer's Guide to Good Practices.* This table provides basic parameters for judging the magnitude of correlation coefficients.

Validity Coefficient Value	Interpretation
Above .35	Very beneficial
.21 - .35	Likely to be useful
.11 - .20	Depends on circumstances
Below .11	Unlikely to be useful

It is critical to note that, depending on the sample size of the study, lesser correlations may be found to be statistically significant. This table is simply intended to be a basic guide.

References

Alreck, P.L. & Settle, R.B. (1995). *The Survey Research Handbook*, 2nd edition. Chicago: Irwin.

Roscoe, J.T. (1975). *Fundamental Research Statistics for the Behavioural Sciences*, 2nd edition. New York: Holt Rinehart & Winston.

U.S. Department of Labor, Employment and Training Administration (1999). *Testing and Assessment, An Employer's Guide to Good Practices.*

Copyright © 2010, IOS, Inc.