ACET inc

Evaluation Take Aways

Free Evaluation Technical Assistance from the Staff of ACET, Inc.

Understanding Measurement Validity

Validity is not new to many people; in everyday examples, something that is considered to be valid tends to be well grounded, logical, and appropriate for the situation. For example, in an argument, one might concede that her opponent has made a valid point or, when frustrated, one might vent to a friend, hoping to have his feelings validated by learning his friend feels similarly. In evaluation, validity of instruments refers to the extent that an instrument or assessment measures what it is supposed to measure. Ensuring an instrument has high validity allows for greater confidence that results from that assessment will be useful, appropriate, and meaningful.

There are many different types of validity in research and evaluation; this *Evaluation Take Away* will focus on four basic types that are common in evaluating instruments, their definitions, and benefits and shortcomings of each of them. While other types of validity, such as external validity, or the extent to which results can be generalized beyond the study sample, are common in research settings, the four validity types discussed here are commonly used in evaluating measurements and are relatively simple to assess. After reading, you should be able to identify and understand these four types of validity, as well as begin to apply these principles to your program's evaluation.

Face Validity

Face validity is the degree to which an instrument appears to measure what it claims. In other words, does an instrument look like it measures what it's supposed to measure?

For instance, let's say a school district wants to survey students to evaluate an after school program. The survey would have face validity if it looked like it measured an after school program. In other words, if the after school survey had items on which students rated their enjoyment of the program and how much they liked the staff. In contrast, if the survey had items about students understanding of microeconomics, the survey would not have face validity.

Although face validity is easy to determine, it has limited usefulness. Specifically, just because an assessment *looks* like it will measure a concept doesn't mean it does measure that concept.

Content Validity

Content validity refers to how well the measurement tool measures the entire scope of a topic. In the case of the after school program evaluation, the survey would not have content validity if the survey only measured students' satisfaction with the snacks served. In order to have content validity, the after school program survey should measure student satisfaction across the entire program. So if the program served snacks and offered homework help the survey should have items for both snacks and homework help.

Content validity requires a careful alignment of the measurement tool with the program. And while it is an improvement over face validity, content validity does require an investment in time from the evaluator, staff, and/or stakeholders to ensure that all aspects of a program are included in the measurement tool.

Types of Validity

Type	Description
Face Validity	Does the instrument *look* like it measures what it is supposed to?
Content Validity	Is the entire scope of the topic being measured? Is anything missing?
Criterion-Related Validity	Is the new instrument related to existing criteria or external variables?
Construct Validity	Does the new instrument accurately measure the construct?

Criterion-Related Validity

Face and content validity both emphasize the appearance or content of the measurement tool. In contrast, criterion-related validity determines the degree that two instruments *or* measures assess the

same thing. For instance, criterion-related validity is often used with intelligence tests: if two intelligence tests are correlated, then they measure the same thing. College entrance exams are another form of criterion-related validity: performing well on the exams in high school has been correlated with success in college.

In order to determine the criterion-related validity of the after school survey the evaluators would need to find another measure of after school satisfaction and correlate that with survey responses. For example, it's logical that if students are satisfied with the after school program they would attend the program often, so one way to establish the criterion-related validity of the survey would be to correlate survey responses with attendance.

Because criterion-related validity relies on statistics it has distinct advantages over face and content validity that only rely upon appearance. The disadvantage with criterion-related validity is that it relies upon a solid, logical relationship between the two measures. Without that solid relationship it's likely that the correlations – and the criterion-related validity – would be poor.

Construct Validity

Construct validity refers to the accuracy of instruments to actually measure what they purport to measure. Do the questions on our satisfaction survey actually gauge student satisfaction with the programming, or do they just tell us that students prefer spending time in the after-school program to watching television at home after school? Because a construct (here, "satisfaction") can be abstract, construct validity can be difficult to measure. Construct validity can be quantified in two ways. Convergent validity exists when an instrument is highly correlated with a similar instrument and discriminant validity exists when an instrument has a low correlation with an unrelated instrument.

Summary

Validity is important in measurement because it ensures that an instrument or assessment is measuring what it is supposed to measure, and capturing an accurate view of a program or construct. Without high levels of validity, results from an instrument may not be meaningful – which means a waste of time and money on behalf of the program and its evaluation.

It is important to note that validity is not a black-and-white designation; an instrument can be "highly valid" or "somewhat valid," or may be more valid in one context over another. In addition, validity is not a permanent status and, as populations or programs change, validity should be reassessed. For example, the student satisfaction survey may be highly valid in the district in which it is currently implemented, but if the student population shifts or the program changes, the satisfaction survey may show a decrease in validity. In addition, the validity of the satisfaction survey could also decrease if the after school program was implemented by staff in a new district. As a result, the validity of an instrument should be assessed on an ongoing basis.

In summation, there are many different types of validity, and there are benefits and shortcomings of each. Additionally, an instrument that is highly valid in one context or for one program may not be considered valid in another situation, or for an extended period of time in the sample population. Continuous collection of evidence demonstrating the validity of an instrument is vital to accurate measurement.

References

NCS Pearson. (1995). The importance of validity: How to avoid garbage in, garbage out. *Research Notes*. Retrieved May 2, 2013, from http://www.reedpetersen.com/portfolio/pe/ncspearson-2/research-notes/95-08.htm

Oswald, K., & Price, P. (2006). Validity in measurement. *Research Methods by Dummies*. Retrieved May 2, 2013, from http://psych.csufresno.edu/psy144/Content/Measurement/validity.html

For additional information about this or other ACET, Inc. resources, or for evaluation assistance, please contact:

ACET, Inc.
info@acetinc.com
952.922.1811

Suggested Citation:
ACET, Inc. (2013). *Understanding measurement validity*. Minneapolis, MN.