

Building Validity and Reliability while Diminishing Bias

Daniel Littleton

The University of Arkansas at Little Rock

LSTE 7325

Abstract

With so many different options there are numerous ways an instructional designer can build an assessment. However, the areas of validity, reliability, and bias can't be ignored if the designer wants to achieve an effective outcome. This paper discusses these three essential assessment instruments particularly when building assessments for healthcare organizations.

Building Validity and Reliability while Diminishing Bias

An assessment tool is only effective if it is evaluating the skill or outcome which it is designed to evaluate, and effective assessment depends on identifying effective items while diminishing the ineffective items. There are a number of ways to reach this goal, but the three key areas of measurement that must be satisfied by an assessment item are validity, reliability and bias. These areas of measurement are especially important when building assessments for healthcare organizations.

Discussion

There are many ways to build assessments for healthcare organizations. Because of the diverse nature of what a hospital looks like, employees can be tested through written tests, simulations, or even orally. However, eLearning can be used to develop skills across all areas of a hospital no matter how diverse. Particularly, eLearning modules can be used to improve employee knowledge and abilities.

In my assessment concept I plan to use eLearning modules to address these needs. However, before these modules can be created a task analysis must be done to obtain the knowledge, skills, attitudes, and abilities needed to achieve optimum performance. Once I have an understanding of the job, only then will a module be created. In this module, I will use the test-retest method by providing the employee with a pre-test followed a scenario based post-test. McAlpine (2002) states “where there is little test-retest reliability the people who did well the first time round may not do well the second time round. Obviously this is an important consideration as it suggests that some element of the measure may be due to chance rather than actual, skills, ability, and knowledge” (p. 12). The results of this test-retest method will be

reported to department heads who will also be involved in future and ongoing performance assessment.

Nonetheless, a task analysis and my knowledge of the subject will not be enough. There are other areas that must be addressed in order for these modules to be successful. In particular, there are three critical elements of any effective assessment; validity, reliability, and bias.

Validity

The first critical element to an effective assessment item is validity. Content validity is often evaluated by examining the plan and procedures used in test construction (Brualdi, 2002). An assessment item is considered valid if it measures the skill in question. These can be extremely difficult to write when there is more than one way to arrive at the answer to a question.

Of course, it is entirely possible that the instructor is unable to recognize that an invalid item is invalid. It is only human and natural psychology: if one is trying to write a task that can be solved by one method, it may not cross the instructor's mind that another approach is possible, particularly when the first method is the most efficient approach.

I plan to approach this problem with careful item creation and by using a panel of supervisors familiar with the material and skills needed to successfully complete the task. According to Fitzpatrick (1983), when a person wishes to use a test to appraise performance of a particular area of interest, the relevance of the test's content to this area will be important to determine. By establishing that the content of a test represents important aspects of the area, a user can gain logical support for the claim that performance on the test is indicative of performance in that area (Fitzpatrick, 1983, p. 8). With the help of the supervisors, my content will be more relevant and related, as Fitzpatrick suggested. The supervisors can examine the

suggested assessment item and decide what that specific item is intended to measure. They can also help lay out all of the criteria that should be met in the program.

Reliability

The second critical element of an effective assessment is reliability. Reliability speaks to reproducibility and refers to the consistency of assessment outcomes. A reliable assessment item is one which will produce the correct response from students of sufficient ability over multiple applications and is a major source of validity evidence for assessments. According to Downing (1994), “Low reliability indicates that large variations in scores can be expected upon retesting. Some of the most important constructs assessed in medical education are concerned with behavior and skills such as communication, history taking, diagnostic problem solving and patient management in the clinical setting” (p. 1009) .

In order to improve the reliability of my assessments, I will try to maximize the number of questions or prompts, aim for middle difficulty questions, make certain that all assessment questions are unambiguous and clearly written, and are critiqued by content-expert reviewers. I will use a simple form of pretesting with the same assembled panel of supervisors for item tryout and will bank recommended items as means of improving reliability.

After I feel I have a good mix of unambiguous clearly written expert reviewed questions, I will be gathering information on stability reliability. If my results come back positive with minimal individual failure then I will only investigate the individual failures. If I get negative feedback as a whole then a different form of assessment may be used all together.

Bias

The last critical element of assessment evident in almost all forms of testing is bias. In order for an assessment to truly be effective then bias must be limited or eliminated altogether. It

occurs when student performance on assessment items varies between two groups of students with equal ability levels. As such, bias skews both the validity and reliability of assessment items, so a deliberate effort needs to be made to reduce bias as much as possible.

To begin, I would address bias by personally trying to "review every item in every assessment from the perspective of whether there is anything present in the item that might offend or unfairly penalize any of my students" (Popham, 2014, p. 137). To achieve the least amount of bias, I must be mindful to select personal experiences that are shared by students of both genders. According to Ross (2008) it is important to recognize that as human beings, our brains make mistakes without us even knowing it. The new science of "unconscious bias" applies to how we perceive other people. We're all biased and becoming aware of our own biases will help us mitigate them in the workplace (Ross, 2008, p.15). Ross's theory will be especially important when looking for signs of possible cultural bias due to the high amount of diversity that exists in the healthcare industry.

Once again, I will be using a small diverse assembly panel of department supervisors for the judgment approach. After the assembly, the meaning of assessment bias would be explained. It will then be up to the supervisors to judgmentally scrutinize the procedure.

Conclusion

Because of the diversity of a healthcare organization, it can be difficult to diminish ineffective items and identify effective ones. However, with proper planning any assessment can be effective by taking into account validity, reliability, and bias. These three areas have to be a valued part of the assessment building process, and training in healthcare should be treated no differently.

References

- Brualdi, A. (2002). Traditional and Modern Concepts of Validity. ERIC/AE Digest. Retrieved from <http://www.ericdigests.org/2000-3/validity.htm> (accessed 1 March 2014).
- Downing, S. M. (2004). Reliability: on the reproducibility of assessment data. *Medical Education*, 38(9), 1006-1012. doi:10.1111/j.1365-2929.2004.01932.x
- Fitzpatrick, A. (1983). The Meaning of Content Validity. *Applied Psychological Measurement*, 7(1), 3-13.
- Popham, J. (2014). *Classroom Assessment*. Boston: Pearson.
- McAlpine, M. (2002). Principles of Assessment (CAA Centre, University of Luton), Retrieved from: <http://caacentre.lboro.ac.uk/dldocs/Bluepaper1.pdf> (accessed 1 March 2014).
- Ross, H. (2008). Exploring Unconscious Bias. *CDO Insights*, 2(5), Retrieved from <http://www.cookcross.com/docs/UnconsciousBias.pdf> (accessed 1 March 2014).