Advances in Medical Education and Practice

Dovepress
open access to scientific and medical research

Open Access Full Text Article

PERSPECTIVES

Four tenets of modern validity theory for medical education assessment and evaluation

This article was published in the following Dove Press journal:
Advances in Medical Education and Practice
3 August 2017
Number of times this article has been viewed

Kenneth D Royal

Department of Clinical Sciences, North Carolina State University, College of Veterinary Medicine, Raleigh, NC, USA

Abstract: Validity is considered by many to be the most important criterion for evaluating a set of scores, yet few agree on what exactly the term means. Since the mid-1800s, scholars have been concerned with the notion of validity, but over time, the term has developed a variety of meanings across academic disciplines and contexts. Accordingly, when scholars with different academic backgrounds, many of whom hold deeply entrenched perspectives about validity conceptualizations, converge in the field of medical education assessment, it is a recipe for confusion. Thus, it is important to work toward a consensus about validity in the context of medical education assessment. Thus, the purpose of this work was to present four fundamental tenets of modern validity theory in an effort to establish a framework for scholars in the field of medical education assessment to follow when conceptualizing validity, interpreting validity evidence, and reporting research findings.

Keywords: health education, psychometric properties, validity evidence, educational measurement, health professions education, reliability

Four tenets of modern validity theory for medical education assessment and evaluation

Validity is considered by many to be the most important criterion for evaluating a set of scores,[1–3] yet few agree on what exactly the term means. Since the mid-1800s, scholars have been concerned with the notion of validity, but over time, the term has developed a variety of meanings across academic disciplines and contexts.[4] Accordingly, when scholars with different academic backgrounds, many of whom hold deeply entrenched perspectives about validity conceptualizations, converge in the field of medical education assessment (broadly defined), it is a recipe for confusion. Thus, developing a consensus meaning in the field of medical education assessment seems unlikely. However, all progress must begin somewhere; thus, the purpose of this work was to present four tenets of modern validity theory that largely have reached consensus in educational and psychological assessments.

The notion of validity is routinely acknowledged in the medical education assessment literature, but the use of the term generally is out of alignment with modern validity theory in the educational and psychological sciences. Only on relatively few occasions do medical education researchers acknowledge theories and positions posited by more modern scholars ("modern" means post 1990) who specialize in validity theory. While it is true that experts in validity theory have not reached consensus about validity conceptualizations (and likely never will), there are four basic tenets that most modern validity theorists will agree. The following is an overview of these tenets.

Correspondence: Kenneth D Royal
Department of Clinical Sciences, North Carolina State University, College of Veterinary Medicine, 1060 William Moore Drive, Raleigh, NC 27609, USA
Tel +1 919 513 6100
Email kdroyal2@ncsu.edu

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/AMEP.S139492

Tenet 1 – validity refers to inferences, not instruments

A quick survey of the medical and medical education research literature will readily identify countless instances in which researchers refer to an instrument, albeit educational or clinical, as being valid. For example, a general PubMed search of the past 5 years (February 2012–February 2017) found 753 instances in which the term valid instrument was used and 2,047 instances in which the phrase valid and reliable instrument was used. Of course, many clinical and educational journals in the health professions are not PubMed indexed, so true estimates of inappropriate terminology usage will likely greatly exceed these numbers.

In the context of educational assessment, validity pertains to the inferences or interpretations made about a set of scores, measures, or other results, as opposed to a property of an instrument.[5,6] The reasoning behind this perspective largely is twofold. First, it is well understood that an instrument administered to two different samples may result in vastly different measures of performance. For example, a biochemistry examination will undoubtedly yield very different scores if administered to a sample of medical students and a sample of elementary school children. Clearly, there is nothing inherently "valid" about the instrument itself, as the instrument is prone to yielding different results depending upon the sample frame assessed. Second, latent traits (e.g., ability, competency, attitudes, etc.) are abstractions and cannot be directly observed. Therefore, researchers at best can only indirectly study these traits by using instruments that are intended to capture the trait(s) in question. Given this dynamic, researchers must make inferences about the findings, and it is the extent to which the inferences are accurate and appropriate that they may be valid (or not). The *Standards*,[1] a joint publication by the American Psychological Association (APA), American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME), states validity refers to the collective evidence (theoretical and empirical) that supports the intended use and interpretation of scores. The Standards explicitly state "It is the interpretations of test scores for proposed uses that are evaluated, not the test itself… it is incorrect to use the unqualified phrase 'the validity of the test'".

Tenet 2 – validity evidence, interpretation, and use

Most medical education papers tend to discuss validity in a fragmented manner. That is, authors will perform a study and evaluate only a particular "type" of validity (e.g., content validity, predictive validity, concurrent validity, external validity, and, most unfortunately, face validity,[7] among others). Again, a quick PubMed search using the same aforementioned parameters found the following number of instances each key term was used: construct validity – 5,829, content validity – 2,567, convergent validity – 2,486, concurrent validity – 1,993, predictive validity – 1,978, discriminant validity – 1,836, external validity – 1243, criterion validity – 1,210, face validity – 1,035, internal validity – 660, divergent validity – 499, criterion-related validity – 375, and conclusion validity – 14.

Validity theorists have articulated that most discrete "types" of validity are ad hoc in nature, whereas "construct validity is the whole of validity from a scientific point of view."[8] Messick[9] attempted to unify validity conceptualizations in the 1980s and 1990s, which was perhaps the closest validity theorists have ever come to reaching a consensus about validity. In recent years, however, Messick's uniformed conceptualization has been contested by several leading validity theorists (e.g., Greg Cizek, Michael Kane, among others) as being too ambitious (e.g., attempting to integrate science and ethics) or too complicated. Nonetheless, most scholars continue to agree that it is advisable to weigh the accumulation of collective evidence in order to determine if an inference is adequately supported. The uniform conceptualization of validity has continued to be adopted by the most recent APA/AERA/NCME standards.

In recent years, the notion that validity arguments are contingent upon the interpretation and use of results has also been accepted by most scholars.[10,11] This perspective not only helps clarify expectations about validity evidence but also places an emphasis on score use, which helps ensure results are interpreted in the appropriate context and used appropriately. The emphasis on specificity typically results in weighing validity evidence in light of the complexity of the attribute of study. For example, Kane has argued that when an attribute being studied is fairly simple and straightforward, it should require only a small but reasonable amount of evidence to substantiate the inference. Likewise, attributes that are more complex will necessitate a greater amount of evidence to substantiate the inference.

Tenet 3 – validity is a continuum

Another common tendency is for medical educators to treat validity as a dichotomy. That is, researchers will often conclude something is either valid or not based simply on the presence or absence of a characteristic. This extreme

view is dangerous and may be encumbered with unintended negative consequences, such as dismissing inferences, that largely are fair and trustworthy and/or suggest a particular road of inquiry leads to a dead end. In truth, validity is a continuum onto which cumulative evidence is weighed and judged to support an inference.[12] Suffice it to say, there are varying degrees of validity evidence and multiple ways in which researchers can construct arguments in support of validity evidence. Even within the same study, it is entirely plausible for a researcher to state that there is a great deal of validity evidence to support (aspect X), but limited validity evidence to support (aspect Y). Furthermore, given that there is some element of error associated with all measurements, researchers need to be particularly careful not to speak in terms of absolutes.

Tenet 4 – validation is an ongoing process

The term validation has a distinct meaning from validity. Whereas validity tends to refer to a conceptual framework for interpreting evidence, validation refers to the practice of incorporating and applying validity theory to evaluate evidence. In most medical education studies, researchers tend to present validity evidence in a static manner. That is, although researchers go to great care to accurately present their findings and responsibly acknowledge any potentially confounding variables or other factors that might influence the interpretation of results, there often remains an assumption that results (within a given margin of error) are permanent. Recognizing the complex nature of latent trait measurement, most validity theorists contend that validation is an ongoing process because multiple factors (e.g., new populations/samples of participants, differing contexts, changing knowledge, increased experience, and so on) are subject to change. In fact, because of the often unstable nature of variables in the human and social sciences, some validity theorists have referred to validation as a "never-ending process."[13] It is for this reason that replication studies and periodically revisiting a research question are especially important in medical education research, as additional studies help researchers understand if phenomena have a tendency to change across samples and/or over time.

Conclusion

The notion of validity means many different things to many different people. Validity theorists in the broader fields of education and psychology have debated the definition of the term for well over a century, and these debates remain ongoing today. At present, the outlook for validity scholars reaching a complete consensus on validity appears grim. However, there are four fundamental tenets that most, but certainly not all, validity scholars will agree. This work attempted to summarize these four basic tenets in an effort to establish a framework for scholars in the field of medical education assessment to follow when conceptualizing validity, interpreting validity evidence, and reporting research findings in this field. Given the critical importance of validity, failure to adopt a common framework will only continue to hinder the growth of the field.[14]

Disclosure

The author reports no conflicts of interest in this work.

References

1. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: APA; 2014.
2. Ebel RL. Must all tests be valid? *Am Psychol*. 1961;16(10):640–647.
3. Koretz D. *Measuring Up: What Educational Testing Really Tells Us*. Cambridge, MA: Harvard University Press; 2008.
4. Newton PE, Shaw SD. *Validity in Educational & Psychological Assessment*. London: SAGE Publications Ltd; 2014.
5. Kane MT. Validation. In: Brennan R, editor. *Educational Measurement*. 4th ed. Westport, CT: Praeger; 2006:17–64.
6. Messick S. Validity. In: Linn RL, editor. *Educational Measurement*. 3rd ed. New York, NY: Macmillan; 1989:13–103.
7. Royal K. 'Face Validity' is not a legitimate type of validity evidence! *Am J Surg*. 2016;212(5):1026–1027.
8. Loevinger J. Objective tests as instruments of psychological theory. *Psychol Rep*. 1957;3:635–694.
9. Moss PA. Themes and variations in validity theory. *Educ Meas*. 1995;14(2):5–13.
10. Kane MT. An argument-based approach to validity. *Psychol Bull*. 1992;112(3):527–535.
11. Kane MT. Validating the interpretation and uses of test scores. In: Lissitz RW, editor. *The Concept of Validity: Revisions, New Directions and Applications*. Charlotte, NC: Information Age Publishing; 2009:39–64.
12. Zumbo BD. Validity: foundational issues and statistical methodology. In: Rao CR, Sinharay S, editors. *Psychometrics*. Amsterdam, Netherlands: Elsevier Science; 2007:45–79.
13. Shepard LA. Evaluating test validity. *Rev Res Educ*. 1993;19:405–450.
14. Royal KD, Rinaldo JCB. There's education, and then there's education in medicine. *J Adv Med Educ Prof*. 2016;4(3):150–154.