Psicothema

Psicothema 2014, Vol. 26, No. 1, 97-99
doi: 10.7334/psicothema2013.255

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Validating assessments: Introduction to the Special Section

Stephen Sireci[1] and José-Luis Padilla[2]
[1] University of Massachusetts Amherst (USA) and [2] University of Granada

Abstract

Background: Validation is the process of providing evidence that tests and questionnaires are adequately and appropriately fulfilling the purposes for which they are developed. In this special issue, experts from several countries describe specific approaches to test validation and provide examples of their approach. These approaches and examples illustrate the validation framework implied by the Standards for Educational and Psychological Testing. **Method:** We describe the Standards' approach for building a validity argument based on validity evidence based on test content, response processes, internal structure, relations to other variables, and testing consequences. **Results:** The five articles provide comprehensive examples of gathering data regarding these five sources of evidence and how they contribute to the validation of the use of test scores for particular purposes. **Conclusions:** These five articles provide concrete examples of how the five sources of validity evidence suggested by the Standards can be used to develop a sound validity argument to support the use of a test for its intended purposes.

Keywords: Testing standards, validity, criterion-related validity, content validity, response processes, test validation.

Resumen

Evaluaciones de validez: introducción a la Sección Especial. **Antecedentes:** la validación es el proceso de aportar evidencias de que las evaluaciones mediante tests y cuestionarios cumplen adecuada y apropiadamente los objetivos para los que se elaboran. En este número especial expertos de varios países describen enfoques específicos para la validación y aportan ejemplos. Estos enfoques y ejemplos ilustran el marco de validación implicado por los Standards for Educational and Psychological Testing. **Método:** describimos la aproximación de los Standards para elaborar un argumento de validez a partir de evidencias de validez basadas en el contenido del test, los procesos de respuesta, la estructura interna, las relaciones con otras variables y las consecuencias del uso del test. **Resultados:** los cinco artículos aportan ejemplos comprehensivos de obtención de datos en relación con las cinco fuentes de evidencia, y de cómo contribuyen a la validación del uso de las puntuaciones en el test para objetivos específicos. **Conclusiones:** los cinco artículos aportan ejemplos concretos de cómo las cinco fuentes de evidencias de validez sugeridas por los Standards pueden utilizarse para elaborar un sólido argumento de validez que apoye el uso de test para sus objetivos previstos.

Palabras clave: examinando los standards, validez, validez relacionada con el criterio, validez de contenido, procesos de respuesta, validación del test.

There are two activities fundamental to educational and psychological assessment—instrument development and validation. Validation refers to the process of gathering, evaluating, and summarizing evidence to support the use of an assessment instrument for its intended purposes. The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999) specify five "sources of evidence that might be used in evaluating a proposed interpretation of test scores for particular purposes" (p. 11). This special issue comprises five articles where experts from across the globe focus on one of these five sources of validity evidence and provide examples of validity studies specific to each source.

Received: August 27, 2013 • Accepted: October 2, 2013
Corresponding author: Stephen Sireci
Faculty of Education
University of Massachusetts Amherst
01003 Amherst (Estados Unidos)
e-mail: sireci@acad.umass.edu

The history of the AERA et al. (1999) *Standards* is relatively long for the young field of psychometrics, dating back approximately 60 years. The first edition was in 1954 (APA, AERA, & National Council on Measurements Used in Education, 1954) and subsequent editions were published in 1966, 1974, and 1985. The description of validity provided in these *Standards* evolved over time to reflect changes in the conceptualization of validity. This evolution is summarized in Table 1. The current framework of the five sources of validity evidence emphasizes that instrument validation involves a comprehensive effort relying on different types of evidence that are part of an integrated body of evidence (described as a "validity argument" by Kane, 1992, 2006, 2013) to support the use of a test for a particular purpose.

At the time of this writing, the current version of the *Standards* has been revised and the new edition is expected to be published soon. Although there are likely to be important updates in the new version, it is already known that the description of validity and the validation framework focused on the five sources of validity evidence will remain. Thus, the articles in this special issue will remain congruent with the new version. It should also be noted that

although the *Standards* are jointly developed by three organizations located in the United States, these organizations include members from across the globe, and as the articles in this special issue illustrate, the *Standards* are influential at an international level.

The five articles that follow describe each of the sources of validity evidence and provide examples of how a validation study in each area can be conducted and how the results of such studies can be used in evaluating the use of a test for a particular purpose. They also illustrate how the results of such studies can contribute to a more comprehensive validity argument. The AERA et al. (1999) *Standards* describe a validity argument as follows,

> A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of test scores for specific uses… Ultimately, the validity of an intended interpretation… relies on all the available evidence relevant to the technical quality of a testing system. This includes evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all examinees… (AERA et al., 1999, p. 17)

Clearly, a single validation study cannot provide a compelling body of evidence to support the use of a test for a particular purpose. However, by understanding all five sources of validity evidence, assessment practitioners can be empowered to conduct research to develop a sound validity argument. The intent of this special issue is to provide such empowerment.

The five articles

In the first article, Sireci and Faulkner-Bond (2014) describe the various conceptualizations of validity evidence based on test content and describe different types of studies that can be conducted to provide such evidence. Due to the particular relevance of this type of evidence to educational testing, the discussion focuses on educational tests, and relates traditional studies of content validity to newer conceptualizations of test-curriculum alignment that are particularly important to mandated educational achievement testing in the United States.

Table 1
Evolution of validity in the Standards

Publication	Validity classifications
Technical recommendations for psychological tests and diagnostic techniques: A preliminary proposal (APA, 1952)	*Categories*: predictive, status, content, congruent
Technical recommendations for psychological tests and diagnostic techniques (APA, 1954)	*Types*: construct, concurrent, predictive, content
Standards for educational and psychological tests and manuals (APA, 1966)	*Types*: criterion-related, construct-related, content-related
Standards for educational and psychological tests (APA, AERA, & NCME, 1974)	*Aspects*: criterion-related, construct-related, content-related
Standards for educational and psychological testing (AERA, APA, & NCME, 1985)	*Categories*: criterion-related, construct-related, content-related
Standards for educational and psychological testing (AERA, APA, & NCME, 1999)	*Sources of evidence*: content, response processes, internal structure, relations to other variables, consequences of testing

In the second article, Rios and Wells (2014) focus on validity evidence based on internal structure. They describe different statistical approaches that can be used to provide such evidence, and focus on confirmatory factor analysis (CFA), which is an increasingly popular technique for evaluating the intended structure of an assessment. In particular, they describe in detail the bifactor model as implemented using a structural equation modeling approach, and illustrate how it can be used to test competing theories of the structure of an assessment. In addition, they also illustrate how multi-group CFA can be used to evaluate the degree to which the structure of an assessment is similar (invariant) across different groups of examinees, test forms, or administration formats. This line of validation research is important in evaluating the fairness of an assessment across subgroups and variations of standardized test administrations.

In the third article, Oren, Kennet-Cohen, Turvall, and Allalouf (2014) focus on validity evidence based on relations to other variables. Their research is specific to a particular test—the Psychometric Entrance Test used for admissions to universities in Israel, and the examples they use to illustrate this type of evidence focus on predictive validity. Specifically, they take a predictive validity approach to evaluate the best means of weighting subscores from this high-stakes admissions test to form composites that enhance the predictive validity of the assessment. Thus, the article illustrates a novel approach to validity evidence based on relations to other variables in contemporary educational assessment.

The fourth article focuses on a relatively new conceptualization of a source of validity—validity evidence based on testing consequences. In this article, Lane (2014) describes an evaluation of testing consequences as a validity argument that focuses on both intended and unintended consequences of a testing program. She draws on theories proposed by Cronbach, Kane, and Shepard to emphasize the important role of consequences in a comprehensive validation endeavor and relates consideration of testing consequences to the "theory of action" underlying a testing program. In addition, she provides instructive examples of validity studies based on the evaluation of testing consequences. Given the relatively young avenue of consideration of testing consequences vis-à-vis a theory of action, this article is likely to become a primary resource for contemporary validity practitioners.

In the fifth and final article, Padilla and Benitez (2014) focus on one of the most difficult sources of validity evidence to gather and analyze—validity evidence based on response processes. They emphasize the importance of this type of evidence in a comprehensive validation endeavor and relate it to other sources of evidence, such as evidence based on test content. They also describe different methods for gathering this type of evidence and provide numerous references to applied studies in this area. In their review, they highlight the use of cognitive interviews for gathering validity evidence based on response processes.

Taken together, these five articles illustrate state-of-the-art approaches to validation that are instructive for contemporary psychometricians. They illustrate not only the comprehensiveness of the different approaches, but also the possibilities. Given the research described in these articles it is clear validity practitioners

cannot use excuses that gathering these different types of validity evidence is too difficult. As the five articles demonstrate, investigating these five sources of validity evidence is not only possible, it is already being done.

We are extremely grateful to the authors for providing these illustrations. This special issue of *Psicothema* represents a monumental contribution to the validity literature—one that is bound to be instructive to us for years to come.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychological Association, Committee on Test Standards (1952). Technical recommendations for psychological tests and diagnostic techniques: A preliminary proposal. *American Psychologist*, 7, 461-465.

American Psychological Association (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin, 51*, (2, supplement).

American Psychological Association (1966). *Standards for educational and psychological tests and manuals*. Washington, D.C.: Author.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (1974). *Standards for educational and psychological tests*. Washington, D.C.: American Psychological Association.

Kane, M.T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*, 527-535.

Kane, M.T. (2006). Validation. In R.L. Brennan (Ed.), *Educational measurement* (4th edition, pp. 17-64). Washington, DC: American Council on Education/Praeger.

Kane, M.T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50(*1), 1-73.

Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema, 26*, 127-135.

Oren, C., Kennet-Cohen, T., Turvall, E., & Allalouf, A. (2014). Demonstrating the validity of three general scores of PET in predicting higher education achievement in Israel. *Psicothema, 26*, 117-126.

Padilla, J.L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema, 26*, 136-144.

Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema, 26*, 108-116.

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema, 26*, 100-107.