



Moving forward: Instruments and opportunities for aligning current practices with testing standards

Jonathan D. Bostic

To cite this article: Jonathan D. Bostic (2017) Moving forward: Instruments and opportunities for aligning current practices with testing standards, *Investigations in Mathematics Learning*, 9:3, 109-110

To link to this article: <http://dx.doi.org/10.1080/19477503.2017.1325662>



Published online: 16 May 2017.



Submit your article to this journal [↗](#)



Article views: 131



View related articles [↗](#)



View Crossmark data [↗](#)

INTRODUCTION TO THE SPECIAL ISSUE



Moving forward: Instruments and opportunities for aligning current practices with testing standards

Jonathan D. Bostic

School of Teaching and Learning, College of Education and Human Development, Bowling Green State University, Bowling Green, Ohio

The focus of this special issue is validity-related issues within mathematics education. I am deeply grateful to the Research Council on Mathematics Learning for supporting this call as well as the associate editors for this special issue: Michele Carney (Boise State University), Erin Krupa (Montclair State University), and Jeff Shih (University of Nevada Las Vegas). Each manuscript was reviewed by an associate editor for the special issue, a mathematics educator with expertise in measurement in mathematics education, a mathematics educator who has reviewed for *Investigations in Mathematics Learning*, and myself. Reviewers provided thoughtful feedback that led to a set of three accepted manuscripts and a 20% acceptance rate. Readers are encouraged to reflect on the purpose, arguments, and evidence within each article, as each uses different approaches, which ultimately lead to appropriate uses and score interpretations.

Instrument quality strongly influences the data collected and relatedly, findings of a research study (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014; Gall, Gall, & Borg, 2007). Instruments with a clearly defined purpose and supporting validity evidence are foundational to conducting high-quality quantitative research (Newcomer, 2012). Near the core of any methodology is the tool used to collect data. The data collected using an instrument is grounded in the validity evidence gathered, and corresponding arguments to support its use in research context. Thus, research aiming to build on past research or lay a foundation for a new vein of quantitative-focused research must be supported by instruments that have diverse and robust validity evidence and arguments.

Validity is a central tenet of effective construct measurement (Messick, 1980). Nearly 30 years ago, Messick (1989) introduced construct validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy and appropriateness of inferences and actions* based on test scores or other modes of assessment” (p. 13, emphasis in original). More recently, Kane (2016) expressed that “validity is a property of the proposed interpretations and uses of the test scores and is not simply a property of the test or of the test score” (p. 64). Validation requires a guiding purpose and evidence supporting a guiding purpose (AERA, APA, & NCME, 2014; Kane, 2016, 2001; Wilson, 2004). Two manuscripts in this special issue address current standards for assessment in educational research and a third manuscript shares a type of validity evidence that might supplement other sources of evidence.

First, Gleason, Livers, and Zelkowski present a validation study of the Mathematics Classroom Observation Protocol for Practices (MCOP²). The authors discussing the MCOP² provide validity evidence that appropriately bounds the use of the instrument and frames score interpretations, which is central to an argument-based approach (AERA, APA, & NCME, 2014; Kane, 2016). The second manuscript, by Eddy, Harrell, and Heitz, describes an observation protocol called AssessToday, which may be used for short-cycle formative assessments. These short-cycle formative assessments have potential for use during day-to-day instruction. The authors provide a discussion about interrater reliability and building a meaningful protocol with results that support appropriate

score uses and interpretations. The third manuscript, by Thompson and Senk, explores a novel type of validity evidence within the validation process: opportunity to learn (OTL). OTL has potential to support content-related validity evidence and, more broadly, appropriately bound generalizations of data gathered from instruments.

In sum, these manuscripts highlight calls for using an appropriate and robust framing for validity according to current evaluation standards (AERA, APA, & NCME, 2014). There is no checklist for validation studies and, more importantly, researchers must critically examine the weakest areas of a measure before selecting it for use, much less drawing conclusions from it (Kane, 2016). Readers are encouraged to reflect on these manuscripts and others as a means to plan, enact, and report high-quality research that stems from measures with sufficient and necessary validity arguments.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Gall, M., Gall, J., & Borg, W. (2007). *Educational research: An introduction* (8th ed.). Boston, MA: Pearson.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38, 319–342. doi:10.1111/jedm.2001.38.issue-4
- Kane, M. T. (2016). Validation strategies: Delineating and validating proposed interpretations and uses of test scores. In S. Lane, M. Raymond, & T. M. Haladyna (Eds.), *Handbook of Test Development* (2nd ed., pp. 64–80). New York, NY: Routledge.
- Messica, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 33–45). Hillsdale, NJ: Erlbaum.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012–1027. doi:10.1037/0003-066X.35.11.1012
- Newcomer, K. (2012). Basics of design for evaluation of cohesion policy interventions. In K. Olejniczak, M. Kozak, & B. Bienias (Eds.), *Evaluating the effects of regional interventions: A look beyond current structural funds practice* (pp. 161–176). Warsaw, Republic of Poland: Ministry of Regional Development.
- Wilson, M. (2004). *Constructing measures: An item response modeling approach*. New York, NY: Routledge.