VALIDATION: A BURGEONING METHODOLOGY FOR MATHEMATICS EDUCATION SCHOLARSHIP

Jonathan Bostic
Bowling Green State University
bosticj@bgsu.edu

Gabriel Matney
Bowling Green State University
gmatney@bgsu.edu

Toni Sondergeld
Drexel University
tas365@drexel.edu

Gregory Stone
Metriks Amerique
gregorystone@metriks.com

Abstract

Validity-related issues are a growing topic within the mathematics education community. Until recently, validation has been treated as something to gather when convenient or is rarely reported in ways that conform to current standards for assessment development. This theoretically-focused proceeding adds to a burgeoning theoretical argument that validation should be considered a methodology within mathematics education scholarship. We connect to design-science research, which is a well-established framework within mathematics education. The goal for this proceeding is to foster the conversation about validation using examples and to communicate information about validation in ways that are broadly accessible.

Introduction

In the last four years, validity issues are taking a greater focus within assessment and measurement of phenomena using quantitative instruments. This is evidenced through a special issue of *Investigations in Mathematics Learning*, National Science Foundation-funded conferences on validity issues within mathematics education contexts, and peer-reviewed manuscripts and books addressing validity and validation issues within the scope of mathematics education scholarship. These scholarly works are springing from mathematics education researchers working collaboratively with others from different disciplines such as learning scientists, psychometricians and research methodologists, and special educators. Grounding ideas in theoretical and methodological frameworks is central to generalizable research that has broader impacts (Confrey, 2018). While there are procedures for validation (e.g., Kane, 2012; Schilling & Hill, 2007), there are few that frame validation as a methodology with its own nuances (e.g., Jacobsen & Borowski, in press). There may be many reasons for why validation has not been treated as a methodology and some of those include but are not limited to (a) pressures not to conduct validation studies, (b) challenges in publishing validation arguments (Bostic, Krupa, Carney, & Shih, in press-a), and (c) decreased emphasis in methodological training of doctoral students in the disciplines (Shih, Reys, Reys, & Engledowl, in press). To that end, this theoretically-focused manuscript aims to augment recent work by Jacobsen and

Borowski (in press) to ground validation work in mathematics education as a methodology akin to design science.

<div align="center">

Relevant Literature

</div>

What is a Methodological Framework?

For this proceeding, we characterize a methodological framework as one that allows a researcher to apply analytical tools to respond to a research question (Cresswell, 2012). For our purposes here, methodology implies ways to conduct research in a manner that synergizes with a chosen theoretical, philosophical, or epistemological framework.

One Approach to Design-science as a Methodology

Design science research was developed to address central questions about learning (Collins, Joseph, & Bielaczyc, 2004). A central component of design research is a "temporal process flowing roughly from conceptualization to realization" (Middleton, Gorard, Taylor, & Bannan-Ritland, 2003, p. 63). Design research can: (a) address theoretical questions about the nature of learning in context, (b) provide a methodological approach for studying learning phenomena in an authentic setting as opposed to laboratory settings, (c) go beyond a singular measure of learning, and (d) derive justifiable findings from formative evaluation (Collins et al., 2004). Design research thus serves scholars as a methodological tool. There are multiple ways to frame design-science methodologies. In sum, a design-science based methodology (e.g., Middleton et al., 2003; Schwartz, Change, & Martin, 2003) fosters "a focus on instruments that both precipitate and measure effects has historically been effective at supporting innovation" (Schwartz et al., 2003, p. 63); in our own research, a test in diverse classroom settings.

One design-science methodological approach has seven phases within its design cycle: (1) grounded models, (2) artifact development, (3) feasibility study, (4) prototying and trials, (5) field study, (6) testing, and (7) dissemination and impact (Middleton et al., 2003). For phase 1, reviews of literature and interfacing with experts helps to ground work on assessment development. It begins to answer questions such as: What will this instrument do? What has already been done in this area of assessment development? How will the interpretations/outcomes from the assessment be used? In phase 2, a rough draft assessment is produced based upon responses to these questions and others. For phase 3, data are gathered to evaluate the quality of the initial draft and make revisions. Cognitive interviews with a measure or real-time observations with an assessment might be used to explore response processes

evidence. In phase 4, revisions are made and a new artifact is produced. A content review committee (i.e., expert panel) or potential typical respondents might then examine the instrument for content, response processes, and/or internal structure validity evidence. In phase 5, implementation studies with a larger sample are conducted to examine the assessment for facets related to internal structure and usability. This sets up for phase 6, when psychometric studies are conducted because there are sufficient (i.e., size and type) data. Finally at phase 7, the developed assessment is disseminated for broad use. This is also the stage where effectiveness studies are conducted to engage questions such as: How sensitive is the assessment to the desired phenomena? Are there quantitative similarities between the assessment and similar instruments? What are the contexts for which might the instrument not be appropriate? Through these seven steps, researchers are able to reify an idea into an actionable product, like an assessment.

Validity and Validation: Definitions and practice

Validity is "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p. 11). Because peer-reviewed manuscripts have historically tied validity to an instrument (see Author, in press-a), it must be re-stated that validity is linked to the interpretations and outcomes - not the assessment. Validity gives scholars confidence that the interpretations from quantitative scores derived from an assessment are the intended ones and not associated with a different construct. The *Standards for Educational and Psychological Testing* (AERA et al., 2014) frame five validity sources for assessment developers and users: test content, response processes, relations to other variables, internal structure, and consequences from testing.

The validation process is cyclical (see Figure 1) in nature and requires iterative loops before an assessment is ready for broad-scale use.
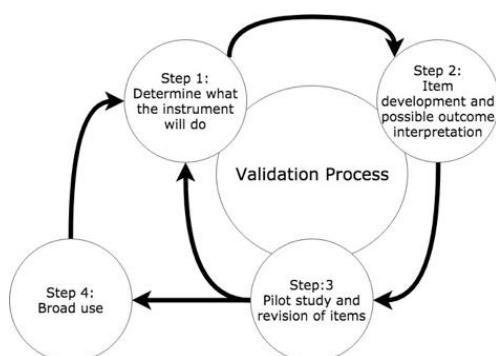
Figure 1. *Validation process. See Gerber, Bostic, & Lavery (2018) for further information.*
The first step is to determine what an assessment will do and what it will measure. This typically requires determining a construct, defined here as "the concept or characteristic that a test is designed to measure" (AERA et al., 2014, p. 11). The second step is developing items and reflecting on ways of interpreting results. During this step, assessment developers think deeply about validity evidence. Drawing across validation frameworks (e.g., Kane, 2012; Schilling & Hill, 2007), this step is likely the most arduous but also the most important. In step three, an assessment is piloted to gather data, inform revisions, and a return to examining the construct that was selected. The reason for returning to step one is that it is possible to move away from the intended construct; therefore, a formative check is warranted. If there is sufficient evidence for the assessment developers suggesting it is functioning adequately, then broader use is acceptable (step 4). Previously, presentations at RCML focused on assessment development have addressed this validation process (e.g., Bostic & Matney, 2018) but not necessarily connected them to validation as a methodology within mathematics education scholarship. Digging into previous work by this research team, Bostic and Matney (2018) presented and fostered discussions at RCML annual meetings around the *Standards* (AERA et al., 2014) and how they were enacted across three assessments ready for broad use in mathematics education scholarship. This proceeding picks up where that one ended and extends work to be more educative and approachable to scholars with a wide range of experience in measurement. In what follows, we connect the validation process, one design-science framework (Middleton et al., 2003), with one problem-solving measure (e.g., PSM6; see Bostic & Sondergeld, 2015) that is a component of a series of measures available for grades 3-8 in Table 1.

It is evident that there are clear connections between validation stages and one design-science framework. Where validation may be a broader term and include many aspects, the design-science framework breaks it down into subcomponents in much the same way sources of validity are categorized in the *Standards* (AERA et al., 2014).

Table 1. Connecting validation and design-science stages with PSMs

Validation	Actions completed in PSM development	Design-science

(1) Determine what the instrument will do	Examine relevant literature, review related assessments, conduct interviews with expert panel	(1) Grounded models
(2) Item development and possible outcome interpretation	Conduct expert panel review, cognitive interviews, small-scale pilot with one class of students	(2) Artifact development
(3) Pilot study and revision of items	Perform small-scale study (approximately 100 respondents), analyze with Rasch modeling, revise items appropriately.	(3) Feasibility study, (4) Prototyping, trials, (5) Field study
(4) Broad use	Perform large-scale study with 300+ respondents.	(6) Testing, (7) Dissemination

A central piece of the validation process is a methodological (i.e., procedural) aim - that is, how to accomplish specific goals. There are specific decisions to be made, which are tied to a desired outcome and chosen theoretical framework (e.g., AERA et al., 2014; Kane, 2012; Schilling & Hill, 2007). These decisions involve when, how, and from whom to collect data - and what manner to analyze those data and for what purpose. Ways to communicate choices for those decisions to potential users is not as simple as a manuscript section labelled participants, instrumentation, data collection, and data analysis. Because the involvement of participants varies at different stages in both design science and validation, it becomes complicated to convey this information. Moreover, the ways information gathered during those stages are analyzed can vary. For instance, assessment developers might choose to analyze a few samples of assessment data at first using one approach and digging deeply into it (e.g., grounded theory; see Charmaz, 2006) whereas later (i.e., broad use) might require a different analytical approach and looking to confirm broad themes through inductive analysis (Hatch, 2002) that were developed earlier in the validation process. Another challenge is that the goal (i.e., assessment being developed) is not validated but its outcomes are. Thus, a central focus on conveying information must be a clear, convincing argument that the outcomes from using an assessment are logically drawn and not that it is merely sound psychometrically.

Current Discussions of Validation as a Methodology

Jacobsen and Borowski (in press) argue that validation acts as a methodological tool that has been underutilized. They and others (e.g., Bostic, 2017, Bostic, Krupa, Carney, & Shih, in press-

a) note the lack of validation work within mathematics education scholarship. Albeit, gathering validity evidence and constructing a validity argument during the design and use phases for an assessment are central to generating generalizable research (AERA et al., 2014; Kane, 2012). Without a validation argument for the interpretations of scores from an assessment, it is uncertain how the scores on that assessment are accurate reflections of an individual or group's attributes (Bostic et al., in press; Kane, 2012). Thus, validation ought to have a central place in mathematics education research that uses quantitative assessments if an aim is to understand factors related to teaching and learning in their authentic settings. Design research draws upon authentic (real world) settings of research and not lab settings. Therefore, validation and design research share a mutual interest in understanding "what is" rather than "what might be".

Implications for Current Assessment Development: A Brief Example

A current National Science Foundation-funded project titled Developing and Evaluating Assessments of Problem Solving (DEAP; NSF #1720646, 1720661) is using the validation stages and a design-science framework (see Middleton et al., 2003) simultaneously to develop a series of measures that assess elementary (i.e., grades 3, 4, and 5) students' problem-solving ability within the context of math content and practices addressed in the Common Core State Standards for Mathematics. This series connects to previously developed measures for grades 6-8. The development team is currently in stage 3 of the validation cycle and is preparing to re-enter the cycle after conducting the initial product and pilot testing. Concomitantly, the team's work might be classified at stage 5 of the design-science framework. More information about current assessment development activities are available (see Bostic, Matney, Sondergeld, & Stone, 2018).

Conclusions and Next Steps

As a result of using validation as a methodology within mathematics education scholarship, assessment developers are better equipped to converse with potential users (e.g., teachers, district representatives, scholars) and those closely associated with test-takers (e.g., students, parents/guardians, school personnel). Data gathering takes a practical approach to inform product development and validate outcomes/interpretations of the assessments. Assessment is central to sound research and without valid outcomes from assessment – the field cannot truly trust their implications. An issue coming from the fervor among mathematics education scholars is that validity must become part of the critical conversation about scholarship that aims to have high

impact (Williams & Latham, 2017). As a result of a growing focus on validity issues, methodological framing of such scholarship becomes a bigger issue. Applying traditional quantitative or qualitative methodologies to communicate scholarship on validity issues and validation arguments presents unnecessary challenges to both authors and readers. Hence, validation should be considered as a viable methodological tool in empirical mathematics education research. We argue that validation as a methodology in mathematics education scholarship has utility. Validation bears striking similarities to design science, which is an established methodology. We recognize that this work and Jacobsen and Borowski (in press) are at the leading edge and more scholarship is needed to better ground validation as a methodology within mathematics education scholarship. Continued validation projects within mathematics education and discussions with diverse scholars will ultimately derive a powerful means for scholars to have broad impact and substantiate intellectual merit for work examining assessment and measurement within mathematics education contexts.

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.

Bostic, J., & Sondergeld, T. (2015). Measuring sixth-grade students' problem solving: Validating an instrument addressing the mathematics Common Core. *School Science and Mathematics Journal, 115, 281-291.*

Bostic, J. (2017). Moving forward: Instruments and opportunities for aligning current practices with testing standards. *Investigations in Mathematics Learning, 9*(3), 109-110. Author.

Bostic, J. & Matney, G. (2018, February). *Current Trends: Improving Test Development and Implementation Practices.* Paper presented at meeting of the Research Council on Mathematics Learning. Baton Rouge, LA.

Bostic, J., Matney, G., Sondergeld, T., & Stone, G. (2018, November). Content validity evidence for new problem-solving measures (PSM3, PSM4, and PSM5). In T. Hodges, G. Roy, & A. Tyminski (Eds.), Proceedings for the 40[h] Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 1641). Greenville, SC.

Bostic, J., Krupa, E., Carney, M., & Shih, J. (in press). Reflecting on the past and thinking ahead in the measurement of students' outcomes. In J. Bostic, E. Krupa, & J. Shih (Eds.), *Quantitative measures of mathematical knowledge.* New York, NY: Routledge.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* Thousand Oaks, CA: Sage.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences, 16*(1), 14-42.

Confrey, J. (2018). Research: To inform, deform, or reform? In J. Cai (Ed.), *Compendium for research in mathematics education (pp. 3-27).* Reston, VA: National Council of Teachers of Mathematics.

Cresswell, J. (2011). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.)* Boston, MA: Pearson.

Gerber, D., Bostic, J., & Lavery, M. (2018, October). *Promoting understanding and applications of validity for teachers' classroom assessment.* Paper presented at the 2nd annual conference on the confluence of classroom assessment and large-scale psychometrics and related disciplines (National Council on Measurement in Education). Lawrence, KS.

Hatch. J.A. (2002). *Doing qualitative research in educational settings.* Albany, NY: State University of New York Press.

Jacobsen, E. & Borowski, R. (in press). Measure Validation as a Research Methodology for Mathematics Education. In J. Bostic, E. Krupa, & J. Shih (Eds.), *Quantitative measures of mathematical knowledge.* New York, NY: Routledge.

Kane, M. T. (2012). Validating score interpretations and uses. *Language Testing, 29*(1), 3-17.

Middleton, J., Gorard, S., Taylor, C., & Bannan-Ritland, B. (2008). The "compleat" design experiment. In A. Kelly, R., Lesh, & J. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics teaching and learning* (pp 21-46). New York, NY: Routledge.

Schwartz, D., Chang, J., & Martin, L. (2008). Instrumentation and innovation in design experiments: Taking the turn towards efficiency. In A. Kelly, R. Lesh, & J. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics teaching and learning* (pp. 47-67). New York, NY: Routledge.

Shih, J., Reys, R., Reys, B., & Engledowl, C. (in press). Examining the career paths of doctorates in mathematics education working in institutions of higher education. *Investigations in Mathematics Learning.* Retrieved from https://doi.org/10.1080/19477503.2018.1514954

Schilling, S. G., & Hill, H. C. (2007). Assessing measures of mathematical knowledge for teaching: A validity argument approach. *Measurement: Interdisciplinary Research and Perspectives*, 5(2-3), 70-80.

Williams, S. & Latham, K. (2017). Journal quality in mathematics education. *Journal for Research in Mathematics Education, 48*(4), 369-396.