EXPLORING AND EXAMINING QUANTITATIVE MEASURES

Jonathan D. Bostic — Bowling Green State U. — bosticj@bgsu.edu
Michele Carney — Boise State U. — michelecarney@boisestate.edu
Erin Krupa — Montclair State U. — krupae@mail.montclair.edu
Jeff Shih — U. of Nevada - Las Vegas — jshih@unlv.nevada.edu

The purpose of this working group is to bring together scholars with an interest in examining the research on quantitative tools and measures for gathering meaningful data, and to spark conversations and collaboration across individuals and groups with an interest in synthesizing the literature on large-scale tools used to measure student- and teacher-related outcomes. While syntheses of measures for use in mathematics education can be found in the literature, few can be described as a comprehensive analysis. The working group session will focus on (1) defining terms identified as critical (e.g., large-scale, quantitative, and validity evidence) for bounding the focus of the group, (2) initial development of a document of available tools and their associated validity evidence, and (3) identification of potential follow-up activities to continue the work to identify tools and developed related synthesis documents (e.g., the formation of sub-groups around potential topics of interest). The efforts of the group will be summarized and extended through both social media tools (e.g., creating a Facebook group) and online collaboration tools (e.g., Google hangouts and documents) to further promote this work.

Keywords: Assessment and Evaluation; Research Methods

Introduction

There is value in the knowledge that large-scale quantitative research can bring to the field in terms of generalizability to educational practice when appropriately conducted (American Statistical Association, 2007; Hill & Shih, 2009). The American Statistical Association's report (2007) on *Use of Statistics in Mathematics Education Research* states:

> If research in mathematics education is to provide an effective influence on practice, it must become more cumulative in nature. New research needs to build on existing research to produce a more coherent body of work… Studies cannot be linked together well unless researchers are consistent in their use of interventions; observation and measurement tools; and techniques of data collection, data analysis, and reporting. (pp. 4-5).

As education has shifted more towards data driven policy and research initiatives in the last 25 years (Carney, Brendefur, Thiede, Hughes, & Sutton, 2016; Hill & Shih, 2009), the data for policy-related aspects are often expected to be quantitative in nature (e.g., end-of-course assessments and numerical value of reform-oriented teaching). Funding agencies encouraging research (i.e., National Science Foundation and Institute of Education Sciences) often request proposals to employ quantitative measures with sufficient validity evidence (see http://ies.ed.gov/ and http://www.nsf.gov/).

Measure (instrument) quality strongly influences the quality of data collected and relatedly, findings of a research study (Gall, Gall, & Borg, 2007). Measures with a clearly defined purpose and supporting validity evidence are foundational to conducting high quality large-scale quantitative work (Newcomer, 2009). There are few syntheses of quantitative tools for mathematics educators to employ and even fewer discussions of the validity evidence necessary to support the use of measures in a particular context. Syntheses of measures for use in mathematics education can be found in the literature but these are typically not intended as a comprehensive analysis. For example, Carney et al. (2015) conducted a brief review of self-report instructional practice survey scales applicable to mathematics education. Boston, Bostic, Lesseig, & Sherman (2015) conducted a review of three widely known classroom observation protocols to assist mathematics educators in determining the appropriate tool for their particular research question and context. Both reviews provided a background on existing measures and their associated validity evidence in relation to a new measure under development. It is important that this type of work continues and is encouraged by the field. Thus, this working group aims to increase conversation around quantitative tools for use on a large-scale with this working group. We share three goals for this proposed working group: (a) To bring together scholars with an interest in examining the research on quantitative tools and measures for gathering meaningful data; (b) To spark conversations and collaboration across individuals and groups with an interest in large-scale tools and those conducting research on student- and teacher-related outcomes; (c) To generate products to disseminate widely across the field of mathematics education scholars.

Related Literature

Historical Context, Terms, and Rationale for Working Group

The National Mathematics Advisory Panel (2008) found that only a "small proportion of those [reviewed] studies have met methodological standards. Most ….failed to meet standards of quality because they do not permit strong inferences about causation or causal mechanisms" (pp. 2-7). Sound methodology is guided by appropriate measure or instrument choice. Good research takes on quantitative, qualitative, and at times both methodologies to become mixed-methodologies (Hill & Shih, 2009; Cresswell, 2012). Our focus for this proposal is quantitative-inclusive methodologies, specifically focusing on measures and tools associated with them, to support mathematics educators use of and need for quantitative tools that may be used in large-scale studies.

Near the core of any methodology is the measure or instrument used to collect data (Newcomer, 2009). The American Psychological Association, National Council on Measurement Education, and American Educational Research Association ([APA, NCME, AERA] 2014; 1999) provide clear guidelines regarding measurement validity and reliability. At a minimum, sufficient evidence for five variables must be shared related to validity: (1) content evidence, (2) evidence for relationship to other variables, (3) evidence from internal structure, (4) evidence from response processes, and (5) evidence from consequences of testing (AERA, APA, & NCME, 1999, 2014; Gall et al., 2007). Unfortunately, "evidence of instrument validity and reliability is woefully lacking" (Ziebarth, Fonger, & Kratky, 2014, p. 115) in the literature. Validation studies of quantitative measures are noticeably absent from mathematics education journals, which present the challenge of determining whether an instrument is appropriate for a given study much less whether it will generate valid and reliable data for analysis (Hill & Shih, 2009). Hill and Shih (2009) reported that eight of 47 studies published in the *Journal for Research in Mathematics Education* provided any evidence related to validity and the majority provided only psychometric evidence. Our goal for this literature review is to present a need for a working group at PME-NA 38 that will bring individuals from around North America to conduct more syntheses and further explore needed areas of tools that can be used to study both student- and teacher- related measures in large-scale research by mathematics educators.

Examining Student-focused Measures

Quantitative measures of student's mathematics content knowledge, problem solving, beliefs, and other factors have been employed across various contexts. We share an initial set of literature to frame the thinking for working group participants. Moreover, we welcome those that have interests not necessarily listed in this section.

Mathematics content knowledge. Students' mathematics content knowledge has been assessed in large-scale studies using end-of-course (high-stakes) measures during the last decade, Trends in Mathematics and Science Study (TIMSS), and National Assessment for Educational Progress (NAEP). Researchers who developed the PISA and NAEP report the validation process; however, the end-of-course measures are often shrouded by commercial entities (e.g., American Institutes of Research and Pearson). The latter group makes examining the quality of the measures for content knowledge problematic. Broadly speaking, it is challenging for researchers aiming to make decisions regarding use of items (or previously used measures) without syntheses describing measure qualities as well as similarities and differences across measures. Thus, a measure may claim to measure students' (at one grade- or developmental-level) content knowledge but how is content knowledge defined for each measure?

Beliefs. Students' beliefs of mathematics, mathematics teaching, and usefulness of mathematics for the real world have been examined in various ways. Students taking the NAEP assessment also responded to questions designed to measure their perceptions of mathematics (Dossey, Mullis, Lindquist, & Chambers, 1988). In the survey created by Dossey and colleagues, students responded to several Likert scale items regarding their attitudes and beliefs about mathematics. Similarly, Lazim, Osman, and Salihin (2004) created a mathematics belief questionnaire that had four belief dimensions: "[about] the nature of mathematics, about the role of teachers, about teaching and learning mathematics, and about their competency in mathematics" (p. 5). Again, the instrument consisted of Likert scale items self-reported by the students. The authors claim they achieved high reliability after the development of the survey but it was not reported. Hence, greater examination of these instruments is needed to benefit mathematics education research.

Examining Teacher-focused Measures

A couple articles have provided syntheses of the literature related to quantitative teacher-focused measures. We explore three sets here: observation protocols (of instruction), teachers' content knowledge, and teachers' beliefs. Again, we use this as a starting point and welcome interests within teacher-focused measures that are not necessarily represented within this frame.

Observation protocols. In 2015, Boston and colleagues compared the Reformed Teaching Observation Protocol, Mathematical Quality of Instruction, and Instructional Quality Assessment. A key finding of the study was that these three unique large-scale teacher-related observation protocols provided three unique lenses into teachers' instruction (Boston et al., 2015). The authors encouraged the field of mathematics education to execute further work to closely examine other observation tools and share syntheses of relevant literature.

Teachers' content knowledge. The components of the Mathematical Knowledge for Teaching (MKT) construct (Ball, Thames, & Phelps, 2008) can serve as a useful tool for exploring and examining quantitative measures of teachers' knowledge. Quantitative measures designed for teacher certification purposes (e.g., the Praxis series) tend to focus on the component of common content knowledge, ignoring other important components of the MKT framework often deemed important to mathematics educators. Other assessments are designed specifically with the intent of measuring teachers' knowledge of particular content areas (e.g., Knowledge of Algebra for Teaching measure, McCrory, Floden, Ferrini-Mundy, Reckase, & Senk, 2012) or grade bands (e.g., Diagnostic Teacher Assessment in Mathematics and Science, Saderholm, Ronau, Brown, & Collins, 2010). The most commonly used quantitative measures for teachers' content knowledge in mathematics come from the Learning Mathematics for Teaching (LMT) project (2005). The LMT assessments aims to measure teachers' content and pedagogical knowledge for teaching and are parsed into different content areas (e.g., K-6 geometry, 6-8 Number and Operations, and 4-8 proportional reasoning; LMT, 2005). A review of the NSF database for measures of teachers' math content knowledge for teaching (a) generating quantitative data, (b) with reliability and validity evidence, and (c) could be used in large-scale studies resulted in 16 measures, 11 of which were part of the set from the LMT series. While tools such as the NSF database or the National Council for Teachers of Mathematics Handbook Chapter "*Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts*" (Hill, Sleep, Lewis, & Ball, 2007) provide a brief summary of some potential measures a mathematics education researcher could use to examine teachers' knowledge, it does not provide a comprehensive synthesis that might aid in determining which measure to use for a given research question, much less describe the validity evidence associated with the measure. Again, there is no available synthesis of available tools to measure teachers' knowledge of mathematics.

Beliefs. Philipp (2007) defines beliefs as "held understandings, premises, or propositions about the world that are thought to be true. …Beliefs, unlike knowledge, may be held with varying degrees of conviction and are not consensual" (p. 259). Beliefs and attitudes are different; they are related and at times have been discussed synonymously in the literature (Philipp, 2007). One of the oldest and still used measures is the Fennema-Sherman Mathematics Attitude scale (see Fennema & Sherman, 1976). This measure uses a Likert-scale to assess respondents' attitudes towards several domains. The study describes four Likert-scale self-report measures and accurately suggests the limited scope of self-report measures with regards to validity evidence. The Integrating Mathematics and Pedagogy (IMAP, 2004; see also Ambrose, Clement, Philipp, & Chauvot, 2004) is a web-based survey with open-ended items. This measure overcame the challenges of Likert scales, the lack of context for an overall score, and that respondents may give an opinion when one is not naturally held (Ambrose et al., 2004). A search of academic journals for measures of mathematics teachers' beliefs provided numerous hits but few are found in mathematics education journals, much less a synthesis of those available with validity and reliability evidence to be used in studies with large data samples. Put simply, no syntheses of measures in this are shared.

Session Organization and Plan for Engagement

The purpose of this working group is to gather individuals across North America interested in synthesizing the literature on quantitative tools in mathematics education that can be used in studies with large samples to examine student- and teacher-related outcomes. When considering the process for conducting a synthesis of quantitative tools and measures, it may be helpful to think of identifying and compiling tools and measures and their associated evidence separately from summarizing and evaluating the quality of the evidence. A synthesis includes both compilation and evaluation. The sequencing of the activities for the purposes of a working group will begin with compilation followed by evaluation in subsequent follow-up activities. It is important for the group to come to consensus on the parameters and frameworks for the synthesis. We recognize that the scope of the working group sessions proposed for PME-NA 2016 must be greatly narrowed. Therefore, we primarily focus on our first two of the three goals for the conference, which are shared here:

1. Bring together scholars with an interest in examining the research on quantitative tools and measures for gathering meaningful data.

2. Spark conversations and collaboration across individuals and groups with an interest in tools for large-scale studies and those conducting research on student- and teacher-related outcomes.

Prior Work

The idea for this working group proposal started at PME-NA 2015. We explored interest across the field from potential attendees before writing this proposal. We sought feedback from colleagues using the Association Mathematics Teacher Educators' (AMTE) bulletin board feature as well as the Service, Teaching, and Research (STaR) list-serv. An interest survey was shared broadly with both groups (i.e., AMTE and STaR members) to gather an idea of the level of interest in this idea. Twenty-six people expressed interest, including from individuals who could not attend AMTE's 2016 annual meeting. We held a follow-up meeting at AMTE to meet with fourteen individuals who expressed interest and were attending AMTE's annual meeting. A majority of those at the AMTE follow-up meeting shared that they planned to attend the working group if accepted for PME-NA 2016. To that end, we plan on organizing the sessions in the following manner to address our two primary goals for the PME-NA 2016 working group session.

Session 1

The first session will begin with introductions, in conjunction with discerning the interests and areas of expertise of those in attendance. This will be followed by a group discussion about the stated purpose and aims of the group and the following guiding questions: (a) What do we mean by the term quantitative tools? (b) What do we mean by the term 'large-scale'? (c) How will we define these terms within the working group? We anticipate this discussion will elicit several additional topics that can be further explored during session 1 and potentially sessions 2 and/or 3. Ideally we will conclude by summarizing the discussion from session 1 including potential definitions for the terms identified as critical (e.g., at-scale, large-scale studies, and quantitative) that will be necessary for bounding the subsequent discussion of currently available tools. At the conclusion of session 1, we will present a tentative framework (see table 1 below) for organizing our subsequent discussions around quantitative tools that can be used with large samples to examine student- and teacher-related outcomes. We will request that session participants return to sessions 2 and 3 with ideas for tools that potentially fit within different areas of the framework.

Session 2.

The second session will begin with a discussion on current perspectives in validity related to the argument-based approach (e.g., Kane 2001, 2016). Finbarr Sloane, an NSF-program officer with expertise in mathematics education, measurement, and evaluation has offered to provide a brief overview and facilitate discussion regarding the argument-based approach to validity. Following Dr. Sloane's presentation, the remaining part of session 2 will involve whole-group discussion around potential measures that address the identified areas using the organizational framework for student- and teacher-related outcomes. A brief overview of the organizational framework will be used to ignite the discussion of specific instruments. Table 1 presents the initial organizational framework that will be presented with the full expectation that the group may modify it during sessions 1 and 2. Group facilitators and attendees may begin by placing some relatively well-known tools within the framework to ensure we have a common understanding of the process.

Table 1: Initial Organizational framework for discussion of measures

	Knowledge	Beliefs	Practice
Teachers			
Students			

Session 3

The third session will primarily focus on placing tools within the organizational framework including any associated citations related to publically available or published validity evidence. Depending upon the size of the group, this work may be conducted in small-groups with a whole-group share-out towards the end of session 3. While a long-term aim is to develop syntheses of the literature related to available tools, we see the primary aim of the working group's meeting at PME-NA 2016 as bringing together individuals interested in this conversation and working together on future collaborative efforts in this area. By the end of the third session, we intend to have an initial draft document of some available tools and their associated validity evidence but we do not anticipate this will be a comprehensive document. We will conclude session 3 with a discussion of anticipated follow-up activities to determine the level of interest and commitment from the group in continuing with this work.

Anticipated Follow-up Activities

As a result of our working group discussion and document development, we anticipate several potential follow-up activities. Participants will greatly influence the specific follow-up activities; however, we outline a potential progression of activities to guide discussion of potential 'next-steps'.

One outcome of the working group sessions is a draft document outlining some of the available tools and their associated validity evidence. An anticipated outcome will be to determine how this document should be further refined and later distributed. This will include explicit discussion of next steps to develop a comprehensive synthesis of the literature for wide dissemination to the mathematics education community.

We see several possible venues for further conversations and work related to developing syntheses of the literature on quantitative tools in mathematics education that can be used with studies of large-scale samples to examine student- and teacher-related outcomes. First, we anticipate using both social media tools (e.g., creating a Facebook group) and online collaboration tools (e.g., Google hangouts and documents) to promote these syntheses. Second, we anticipate using mathematics education conferences venues to further the conversations and synthesis work around the project. More specifically, we plan on proposing to continue the PME-NA working group at the 2017 conference. In addition, we anticipate submitting for a symposium at either the 2017 or 2018 conference of the Association of Mathematics Teacher Educators. Lastly, there is potential to apply for grant funding through a NSF CORE Research proposal to support a conference with a focused outcome of a monograph synthesizing the research literature within a particular area.

References

Ambrose, R., Clement, L., Philipp, R., & Chauvot, J. (2004). Assessing prospective elementary school teachers beliefs about mathematics and mathematics learning: Rationale and development of a constructed-response-format beliefs survey. *School Science and Mathematics, 104*, 56-69.

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. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Statistical Association. (2007). Using statistics effectively in mathematics education research. *Retrieved* from http://www.amstat.org/education/pdfs/UsingStatisticsEffectivelyinMathEdResearch.pdf

Ball, D. L., Thames, M. H., & Phelps, G. C. (2008). Content knowledge for teaching: what makes is special? *Journal of Teacher Education, 59*(5), 389-407.

Boston, M., Bostic, J., Lesseig, K., & Sherman, M. (2015). A comparison of mathematics classroom observation protocols. *Mathematics Teacher Educator, 3*(2), 154-175.

Carney, M. B., Brendefur, J. L., Hughes, G. R., & Thiede, K. (2015). Developing a mathematics instructional practice survey: Considerations and evidence. *Mathematics Teacher Educator 4*(1), 1-26.

Carney, M. B., Brendefur, J. L., Thiede, K., Hughes, G., & Sutton, J. (2016). Statewide Mathematics Professional Development: Teacher Knowledge, Self-Efficacy, and Beliefs. Educational Policy, 30(4), 539-572.

Cresswell, J. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.)*. Boston: Pearson.

Dossey, J. A., Mullis, I. V. S., Lindquist, M. M., & Chambers, D. L. (1988). *The mathematics report card. Are we measuring up? Trends and achievement based on the 1986 national assessment.* Princeton, NJ: Educational Testing Service.

Fennema, E., & Sherman, J. (1976). Fennema-Sherman Mathematics Attitude Scales. *JSAS: Catalog of Selected Documents in Psychology, 6*(10). (Ms. No. 1225)

Gall, M., Gall, J., & Borg, W. (2007). *Educational research: An introduction (8th ed.).* Boston: Pearson.

Hill, H. C., Sleep, L., Lewis, J. M., & Ball, D. L. (2007). *Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts.* In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 111-155). Charlotte, NC: Information Age Publishing.

Hill, H., & Shih, J. (2009). Examining the quality of statistical mathematics education research. *Journal of Research in Mathematics Education, 40*(3), 241-250.

Integrating Mathematics and Pedagogy. (2004). *IMAP web-based beliefs-survey manual*. Retrieved from http://www.sci.sdsu.edu/CRMSE/IMAP/main.html

Kane, M. T. (2001), Current concerns in validity theory. *Journal of Educational Measurement, 38*, 319–342.

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.

Lazim, M. A., Osman, M. T., & Salihin, W. A. (2004). The statistical evidence in describing the students' beliefs about mathematics. *International Journal for Mathematics Teaching and Learning, 6*(1), 1-12.

Learning Mathematics For Teaching. (2005). *Mathematical Knowledge for Teaching Measures*. Ann Arbor, MI.

McCrory, R., Floden, R., Ferrini-Mundy, J., Reckase, M. D., & Senk, S. L. (2012). Knowledge of algebra for teaching: A framework of knowledge and practices. *Journal for Research in Mathematics Education, 43*(5), 584-615.

National Math Advisory Panel. (2008). *Foundations for success the final report of the National Mathematics Advisory Panel.* Washington, D.C.: U.S. Dept. of Education.

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). Republic of Poland: Ministry of Regional Development.

Philipp, R. (2007). Mathematics teachers' beliefs and affect. In F. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257-315). Charlotte, NC: Information Age Publishing.

Saderholm, J., Ronau, R., Brown, E. T., & Collins, G. (2010). Validation of the diagnostic teacher assessment of mathematics and science (DTAMS) instrument. *School Science and Mathematics, 110*(4), 180-192.

Ziebarth, S., Fonger, N., & Kratky, J. (2014). Instruments for studying the enacted mathematics curriculum. In D. Thompson, & Z. Usiskin (Eds.), *Enacted mathematics curriculum: A conceptual framework and needs* (pp. 97-

120). Charlotte, NC: Information Age Publishing.

Acknowledgement: We are grateful for support from NSF award #1644314.