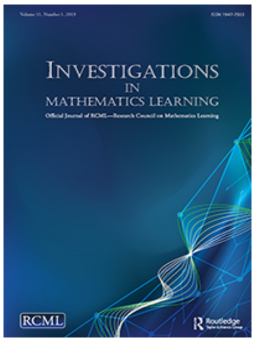

Routledge
Taylor & Francis Group

Investigations in Mathematics Learning

ISSN: 1947-7503 (Print) 2472-7466 (Online) Journal homepage: http://www.tandfonline.com/loi/uiml20

A validation process for observation protocols: Using the *Revised SMPs Look-for Protocol* as a lens on teachers' promotion of the standards

Jonathan D. Bostic, Gabriel T. Matney & Toni A. Sondergeld

To cite this article: Jonathan D. Bostic, Gabriel T. Matney & Toni A. Sondergeld (2019) A validation process for observation protocols: Using the *Revised SMPs Look-for Protocol* as a lens on teachers' promotion of the standards, Investigations in Mathematics Learning, 11:1, 69-82, DOI: 10.1080/19477503.2017.1379894

To link to this article: https://doi.org/10.1080/19477503.2017.1379894

Published online: 25 Oct 2017.

Submit your article to this journal

Article views: 95

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=uiml20

INVESTIGATIONS IN MATHEMATICS LEARNING
2019, VOL. 11, NO. 1, 69–82
https://doi.org/10.1080/19477503.2017.1379894

Check for updates

A validation process for observation protocols: Using the *Revised SMPs Look-for Protocol* as a lens on teachers' promotion of the standards

Jonathan D. Bostic [a], Gabriel T. Matney[a], and Toni A. Sondergeld[b]

[a]Bowling Green State University, Bowling Green, Ohio; [b]Drexel University, Drexel, Pennsylvania

ABSTRACT
The Standards for Mathematical Practice (SMPs) describe mathematical behaviors and habits that students should express during mathematics instruction. Thus teachers should promote them during classroom-based mathematics instruction. The purpose of this article is to discuss the validation process for an observation protocol called the *Revised SMPs Look-for Protocol*. An implication of this study is that users with a robust understanding of the SMPs may feel confident using the protocol as a validated and reliable tool in research and school-based settings. We discuss opportunities and challenges for mathematics teacher educators engaging in classroom observations, in light of this observation protocol.

KEYWORDS
K-12; observation protocol; standards; teachers; validation

A lens on teachers' promotion of the Standards for Mathematical Practice[1,2]

As of 2015, 42 of 50 states within the United States have adopted the Common Core State Standards for Mathematics (CCSSM) as their mathematics standards. The CCSSM has Standards for Mathematics Content and Standards for Mathematical Practice (SMPs; Common Core State Standards Initiative [CCSSI], 2010). SMPs are descriptions of mathematical habits and behaviors and are deeply connected to the National Council of Teachers of Mathematics' process standards (Kanold & Larson, 2012; Koestler, Felton, Bieda, & Otten, 2013). While the CCSSM have been in place for nearly five years, teachers are still struggling to make sense of them, especially the SMPs (Bostic & Matney, 2014). At times, it is unclear to teachers and observers what the SMPs look like during classroom mathematics instruction (Bostic, 2015; Bostic & Matney, 2014). For example, *modeling with mathematics* has a meaning distinct from *modeling as representation* discussed in the K–5 content standards and *mathematical modeling* as described in the high school content standards (Bostic & Matney, 2016; Bostic, 2015). As such, education stakeholders may benefit from having a tool to generate feedback about the ways mathematics teachers' instruction promotes the SMPs. The aim of this article is to present validity evidence for a tool focused on teachers' instruction related to the SMPs and place the tool among others that may act as formative assessments meant to foster teachers' growth. This tool is called the *Revised SMPs Look-for Protocol*. Through this article, mathematics teacher educators will learn about the

CONTACT Jonathan D. Bostic bosticj@bgsu.edu 529 Education Building, Bowling Green, OH 43403.

This article is not currently in submission to another journal. I have gone through the appropriate process of obtaining approval for this research study. I attest that all the authors listed contributed to this article and that no other persons contributed.
[1]This manuscript is supported by multiple grants from two funding agencies. Any opinions expressed herein are those of the authors and do not necessarily represent the views of these agencies.
[2]The authors are especially grateful to Skip Fennell, Elizabeth Kobett, and Jon Wray for their ideas and work on the initial version of the Standards for Mathematical Practice Look-for Protocol as well as the Revised Standards for Mathematical Practice Look-for Protocol.

© 2017 Research Council on Mathematics Learning

nature of developing an observation protocol, negotiating the process of coming to agreement among a coding team, and validity evidence that supports the use of this protocol to evaluate K–12 mathematics teachers' instruction through one lens.

Literature review

Prior validated tools for examining classroom instruction

There are various tools to examine mathematics instruction, which provide insight beyond students' grades and high-stakes test results (Kane & Staiger, 2012; Ziebarth, Fonger, & Kratky, 2014). Observation tools support insight into the teaching–learning dynamic that affects students' learning; they can supplement students' scores. Hence observation tools have great importance for examining teachers' instruction and fostering teachers' growth, with an aim of improving students' outcomes (Bostin, Bostic, Lesseig, & Sherman, 2015; Lesseig, Bostic, Sherman, & Boston, 2017; Kane & Staiger, 2012).

Boston et al. (2015) discusses the strengths and limitations of three validated tools that are often used in educational research for mathematics (i.e., Reformed Teaching Observation Protocol [RTOP], Instructional Quality Assessment [IQA], and Mathematical Quality of Instruction [MQI]). The RTOP was validated for K–16 mathematics and science instruction in the field as well as with videotaped instruction and is meant to focus on reform-oriented teaching (Sawada et al., 2000). A strength of the RTOP is that it provides rich descriptive data and can show changes over time. However, a limitation is that it is not inherently mathematical per se. The IQA also focused on reform-oriented instruction and, additionally, the implementation and discussions surrounding cognitively challenging tasks (Boston & Wolf, 2006; Matsumara, Garnier, Slater, & Boston, 2008). A strength of the IQA is well-defined score levels, while a limitation is that training is substantial, often requiring a couple of days . Finally, the MQI bears similarities to the IQA in that it aims to assess the rigor and richness of mathematics content expressed during classroom instruction (Hill et al., 2008). A strength of the MQI is that it offers users the opportunity to make school-, district-, or scale-level decisions; however, it is limited to use in videotaped observations in K–9 mathematics contexts. The RTOP, IQA, and MQI are all validated observation protocols, yet none of them was intended for use in exploring teachers' promotion of the SMPs. A challenge within classroom observation research is a tool that is effective for capturing evidence of observable teacher behaviors related to the SMPs. That is, if students are expected to engage in the SMPs, then it stands to reason that teachers must promote them. However, how might one connect a teachers' observable action with an SMP, much less do so with a sound validity argument? With the large number of teachers who are working to promote the SMPs among their students, the need for a validated observation tool related directly to the SMPs is warranted.

The *Mathematics Classroom Observation Protocol for Practices (MCOP^2)* responds to the need for a tool that examines the SMPs (Gleason, Livers, & Zelkowski, 2017). "Each of the items on the *MCOP^2* was designed to coordinate with a Standard for Mathematical Practice... for instance, item #9 on the protocol is 'The lesson provided opportunities to examine elements of abstraction (symbolic notation, patterns, generalizations, conjectures, etc.), matching the second Standard for Mathematical Practice that instructors should be aiming to teach their students" (Gleason & Cofer, 2014, p. 96). The *MCOP^2* moves the field forward with a validated tool to examine classroom instruction for the SMPs; however, it has limitations. First, there are only one or two indicators related to some SMPs (e.g., SMPs #2, 4, 5, and 6), hence it is possible to miss potential evidence related to these SMPs. Second, it blends student engagement and teacher facilitation, which may make it difficult to parse how teachers promote the SMPs and students engage in them. Thus, there still exists a need for a validated observation protocol related to K–12 teachers' promotion of SMP-focused instruction that does so in a robust fashion.

Development of the Standards for Mathematical Practice Look-for Protocol

A year after the large-scale adoption of the CCSSM, Fennell, Kobett, and Wray (2013) created a tool called the Standards for Mathematical Practice Look-for Protocol (*SMP Look-for Protocol*). Their goal was to develop and share a tool to gather evidence related to K–12 mathematics teachers' promotion of the SMPs during classroom instruction. An initial version had only one indicator related to each SMP for teachers' promotion of the SMPs and students' engagement in the SMPs. Later versions included observable mathematical behaviors and habits (as many as eight), related to both teachers' and students' observable mathematical behaviors and habits. A final version of the *SMP Look-for Protocol* was shared at the 2013 Association of Mathematics Teacher Educators' annual meeting. Fennell and his team conducted nearly 300 observations and asked numerous mathematics teacher educators, curriculum coaches, and teachers to examine the protocol for their ideas related to it. Synthesizing across groups' voices, it was clear that the protocol was helpful to examine K–12 teachers' promotion of the SMPs during classroom instruction. Fennell and colleagues further shared openness to additional revisions of the protocol. Moreover, they expressed a need to explore validity evidence with the tool in real-time or video-recorded K–12 classroom mathematics observations. The purpose of the present study is to share information related to validating an observation tool for the purpose of analyzing K–12 teachers' promotion of mathematical behaviors and habits framed by the CCSSM SMPs. Our research question is "What validity and reliability evidence supports use of the *Revised SMPs Look-for Protocol* as a tool to examine teachers' mathematics instruction related to promotion of the SMPs?" This protocol has potential to support mathematics teacher educators' formative assessment of K–12 in-service teachers' instruction. A second implication is as a tool in research and evaluation of K–12 in-service teachers' instruction.

Method

Context

We used the protocol developed by Fennell et al. (2013) for approximately 60 observations and were able to chart teachers' growth in promotion of the SMPs through this tool. Through its prior use, our research team felt it was missing some elements related to teachers' promotion of the SMPs. For instance, related to an indicator for SMP #1, "Make sense of problems and persevere in solving them" (CCSSI, 2010, p. 5), we had concerns that counting opportunities where teachers gave students problems with multiple solutions might limit the richness of worthwhile tasks that allowed problem solvers to employ multiple strategies. To that end, we reflected on use of the tool, revised it, and gathered validity evidence for the present observation protocol in this manuscript. Through the stages of the validation, we added some observable aspects related to the SMPs and modified some aspects to better capture teachers' instruction that promoted the SMPs. It is this *Revised SMPs Look-for Protocol* that we explore in our study.

Instrument validation process

The validation process is essential: It gives users of a tool a high degree of certainty that (a) the tool accurately and appropriately captures data related to a known construct (i.e., validity), and (b) the tool consistently measures the construct of interest (i.e., reliability). One way to conduct a validation study for an observation protocol is through an eight-stage process (Artino, La Rochelle, Dezee, & Gehlbach, 2014; Smith, Jones, Gilbert, & Wieman, 2013). The first stage is to conduct a literature review related to the construct of interest. Much like any research, this entails reviewing relevant literature and synthesizing it. The second stage includes conducting interviews and focus groups with content experts and potential users of the protocol to gather more ideas for items. This is important because it begins to establish content validity. Content validity evidence indicates the degree to which content (or items) found on a tool addresses the construct of interest (American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA, APA, & NCME], 2014; Bostic & Sondergeld, 2015; Gall, Gall, & Borg, 2007). Stage three is synthesizing data from literature review, interviews, and focus groups as a means to reach the fourth stage. The fourth stage is item development. For an observation protocol, this entails creating and modifying indicators related to the construct of interest, which are known to potential users. After items are constructed and ready for review, then the fifth stage begins: expert panel validation. This is a step in validation of content for new and/or revised items. Content experts for an observation protocol related to K–12 mathematics education might include mathematics teachers, professional developers, and mathematics education researchers. The sixth stage is to conduct cognitive interviews with potential users of the protocol. Results from this stage highlight areas of strength and weakness, which allow protocol developers the chance to further revise the instrument for ease of use while maintaining content and response processes validity. This stage is iterative and may occur several times until the expert panel, potential users, and protocol developers are pleased with the tool. Once the instrument has undergone these initial phases of development and revision, the protocol is ready for the seventh stage: pilot testing. Pilot testing may highlight any consequences from use. At this point, it is ideal for those testing the instrument to maintain a high degree of interrater agreement or reliability when implementing so that interpretations of results are valid (Crocker & Algina, 2006). That is, analysis of data without sufficient interrater agreement (i.e., $x < 0.9$; James, Demaree, & Wolf, 1993) or reliability (i.e., $x < 0.8$; Gall et al., 2007) leads to spurious conclusions because of a lack of internal consistency. The eighth and final stage is conducting psychometric analysis using data from the protocol (e.g., reliability analysis), and is a means to investigate internal consistency. We included these eight stages in our frame for this validation study of the *Revised SMPs Look-for Protocol*. Table 1 provides a description of each stage as well as a summary of the actions completed during that stage by the authors. We believe that communicating our experiences during each process has promise to inform the work of mathematics teacher educators. That is, validity evidence and experiences using the tool have potential implications for use by others across a variety of mathematics teacher educator contexts.

Table 1. Brief description of stages for validating protocols.

Stage #	Description of stage	Actions completed during this study
1	Literature review	Examined initial protocol, reviewed literature on observation protocols and mathematical proficiency
2	Conducting interviews with content experts and potential users of tool	Conducted interviews with expert panel consisting of K–12 math teachers, math coaches, curriculum coordinators, mathematicians, and mathematics teacher educators.
3	Synthesizing data from literature review and interviews	Employed inductive analysis to generate themes that should be found in indicators
4	Item development	Added and modified indicators from the Fennell et al. protocol
5	Expert panel validation	Submitted revised look-for protocol to expert panel
6	Conducting interviews with potential users of tool and synthesize from these data	Conducted 1–1 and small-group interviews with K–12 math teachers, a principal, curriculum coach, mathematicians performing observations in schools, and mathematics educators
7	Pilot testing tool	Performed 288 observations with the revised look-for protocol
8	Conducting psychometric analysis of data from use of the tool	Performed reliability analysis using Cronbach's alpha (internal consistency) and test–retest reliability

Procedures and analysis for validation

Stage one

We began by exploring the content of the initial protocol developed by Fennell et al. (2013). In our work to revise the protocol, we reexamined the literature for new observation protocols and ideas related to the SMPs. We specifically targeted research since 2010, including manuscripts, proceedings, and presentations. Relatedly, we explored materials published in association with nationally known groups that have a long-standing history in mathematics and statistics education (e.g., National Council of Teachers of Mathematics, National Council of Supervisors of Mathematics, and American Statistical Association). Books such as *Principles to Action: Ensuring Mathematical Success for All* (NCTM, 2014) and the *Guidelines for Assessment and Instruction in Statistics Education* (Franklin et al., 2007) provide suggestions for the critical ideas about mathematics and statistics instruction, specifically, what it means to promote mathematical and statistical behaviors and habits, were consulted. Ideas from the literature supported data gathered in the next stage.

Stage two

Data were collected from an expert panel consisting of 29 individuals from five groups: K–12 mathematics teachers, mathematics coaches, curriculum coordinators, mathematicians (including mathematics instructors teaching mathematics education courses), and mathematics teacher educators who have led professional development focused on the SMPs. K–12 mathematics teachers had one year up to 28 years teaching mathematics. We had at least one teacher from each grade level on the expert panel, for a total of 18 teachers. We reached out to mathematics coaches and curriculum coordinators who worked at a district or school level and provided support to mathematics teachers. In total, three coaches and coordinators were involved at this stage. We sought feedback from two mathematicians with knowledge of the SMPs. Both taught a variety of courses for future K–12 mathematics teachers and had classroom teaching experience. One of the mathematicians held a Ph.D. and the other held a master's degree. Finally, six mathematics teacher educators from across the United States who represented different geographic locations and expertise (e.g., working with elementary, middle, and/or high school teachers) served on the panel. These mathematics teacher educators held a Ph.D. or other terminal degree. The initial developers of the protocol (i.e., Fennell, Kobett, and Wray) agreed to share ideas related to the *Revised SMPs Look-for Protocol* and served in this capacity. Some of the comments from the panel expressed wonder about how we might address the notion of structure in teaching statistics, differentiating modeling with mathematics-focused instruction (see SMP4) from instruction using word problems, and clearly identifying what it means for a teacher to promote strategic use of mathematical tools. For stage two, we communicated with members of the expert panel via email as well as face to face and phone exchanges to make sense of their ideas for a possible tool to gather data about K–12 teachers' promotion of the SMPs during instruction.

Stage three

We synthesized across the data from the literature review and expert panel using inductive analysis. Inductive analysis (Hatch, 2002) allowed the authors to draw impressions that might support item development. Inductive analysis supports users to identify salient themes from datasets (Glaser & Strauss, 1967/2012; Hatch, 2002). Our approach to inductive analysis started with rereading (or relistening) to materials (e.g., expert panel written reviews). Step two was to make memos consisting of initial ideas stemming from this examination of the data. Step three was to reflect on those memos as a way to synthesize them into key impressions, needed as evidence for validity. Step four was to search for evidence within the datasets to support our key impressions. Step five was to search the data for counterevidence. Impressions with a paucity of counterevidence and a large set of evidence were retained. The sixth and final step was crafting clearly written impressions (themes) to keep in mind during item development. One theme drawn from analysis at this stage was that structure in statistics (i.e., the shape, center, and spread of data) was not adequately captured as an indicator of SMP7 by the look-for protocol; indicating a need for revision.

Stage four

The authors drew on impressions generated from the inductive analysis performed in stage three as a means to revise Fennell et al.'s (2013) protocol. No items were deleted during this stage. The authors added and modified indicators from the Fennell et al. (2013) protocol. One example of an addition was indicator 7d, "Encouraging examinations of a 'signal' and 'noise' in statistics-related tasks." As shared in step three, structure in statistics was not addressed in previous versions of the protocol. One example of a modification was the adding of the phrase "and/or strategy" to indicator 1b and 8a. In both cases, instruction drawing on tasks that might have multiple strategies but not necessarily multiple solutions have potential to engage students in mathematical behaviors. A second example and different type of modification included addition and revision. Formerly, Fennell et al. (2013) condensed contextualization and decontextualization for SMP2 as "Provide opportunities for students to decontextualize (abstract a situation) and/or contextualize (identify referents for symbols involved) the mathematics they are learning. We decided to separate this into two related codes as seen in SMP2, "C1. Provide opportunities for students to decontextualize (abstract a situation) the mathematics within a mathematics task. C2. Provide opportunities for students to contextualize (identify referents for symbols involved) the mathematics within a mathematics task." The additions, especially mentioning referents for symbols involved, were needed because it clarified SMP2 in a way that was closely linked to contextualization.

Developing indicators for an observation protocol requires reflection on the intended construct and how it might appear within varied and diverse learning environments, choosing language meant to communicate what to look for during instruction, and intending to create indicators that are specific to one idea. In this instance, the construct was teachers' instructional practices related to the SMPs. Each idea was one SMP (e.g., SMP1, SMP2, etc.).

Stage five

The expert panel reviewed the protocol developed from the previous stage. A total of nine members served on this panel. It consisted of a representative convenience sample of individuals from each group described earlier in stage two (i.e., K–12 teachers, mathematics coaches, curriculum coordinators, mathematicians, and mathematics teacher educators). That is, two teachers from grade bands representative of the licensure areas (i.e., K–3, 4–8; 9–12) volunteered, two mathematics educators, and one mathematician served in this capacity. The panel examined the *Revised SMPs Look-for Protocol* and reflected on the degree to which our revisions and previous statements adequately met the descriptions in the SMPs. We asked them to provide verbal or written feedback indicating whether and/or to what degree the indicators addressed the SMPs. Their feedback was positive, and some suggested clarifying what a specific indicator meant. This feedback was added to the data collected during stage six.

Stage six

Small-group and one-on-one interviews were conducted with those interested in using this tool. A total of eight individuals participated during this stage of the process. Interviews were held with mathematics teachers in leadership positions (i.e., school-wide observers for teacher evaluation): one principal who had previously served as a high-school mathematics teacher, one curriculum coach, two mathematicians, and four mathematics educators currently performing in-school observations related to a grant-funded professional development (PD) project. The goal of these interviews was to learn about the protocol's ease of use and its overall ability to meet the aim of gathering data about K–12 teachers' promotion of the SMPs during classroom mathematics instruction (i.e., addressing content and response processes validity). Inductive analysis was performed on their responses as well as those from the expert panel (stage five) in the same fashion as in stage three. This led to development of the *Revised SMPs Look-for Protocol.*

Stage seven

We piloted the *Revised SMPs Look-for Protocol* during the seventh stage. Data for this validation study came from two sources. The first source consisted of video-recorded data from K–12 teachers located in a Midwest state that adopted the CCSSM. They participated in one of nine grant-funded mathematics PD programs that lasted a minimum of 100 face-to-face hours during one academic year. An objective of these PD programs was to foster teachers' sense making of the SMPs so that they might promote them more effectively during classroom mathematics instruction. Teachers consented to providing videos of instruction prior to the PD and again after 80 hours of PD. The second data source consisted of observations of live instruction in K–12 classrooms conducted by the authors of this article. In sum, the sample included 288 observations of teachers' instruction, which were coded using the *Revised SMPs Look-for Protocol.* Thirty of the 288 observations were made during live instruction, while the other 258 were made using videotaped data (i.e., 129 pre- and 129 post-PD videos). Table 2 describes demographics of schools and teachers in the spirit of describing diversity within the sample. School types were taken from the Ohio Department of Education typology (Ohio Department of Education, 2013). Most teachers were female (84%), working in elementary schools (70%), and working in either rural (40%) or urban districts (32%).

Interrater agreement was high across coders (93%), which exceeds the minimum threshold (90%) needed to conduct reliability analyses (James et al., 1993). Moreover, interrater agreement is stronger than interrater reliability because agreement indicates that coders can be thought of as prescribing codes identically rather than coding in reliably similar patterns (Gall et al., 2007).

Stage eight

Psychometric analysis was conducted during the eighth stage of the validation study to examine internal consistency (reliability) associated with using this tool, in two ways. The first way was internal consistency of the protocol using Cronbach's alpha, which indicates the "coefficient of precision from a set of real test scores" (Crocker & Algina, 2006, p. 117). The second way was test–retest reliability using data from pre- and post-PD observations. A bivariate correlation was used to determine the relationship between pre- and post-PD observations, with higher positive relationships indicating a higher level of test–retest reliability. We addressed two forms of reliability evidence through quantitative analysis in this study. This approach is a primary

Table 2. Demographic data regarding school districts and teachers.

	School district classification				School type			Teacher's sex	
	Urban	Suburban	Small town	Rural	Elementary	Middle	High	Male	Female
Frequency (% of sample)	41 (0.32)	31 (0.24)	5 (0.04)	52 (0.40)	99 (0.70)	20 (0.16)	10 (0.08)	21 (0.16)	108 (0.84)

fundamental assessment of reliability and meets current standards for measurement (AERA, APA, & NCME, 2014).

Instrumentation

The *Revised SMPs Look-for Protocol* includes three or four observable behaviors related to teachers' promotion of each SMP as well as specific notes for observers. The full protocol is shared in Figure 1. Our belief is that teachers' instruction might exemplify some indicators, and possibly one or more indicators across all eight SMPs. It is not anticipated that any teacher's instruction includes marks for every indicator (i.e., 28).

Mathematical Practices	Teachers
1. Make sense of problems and persevere in solving them	☐ A. Involve students in rich problem-based tasks that encourage them to persevere in order to reach a solution ☐ B. Provide opportunities for students to solve problems that have multiple solutions. ☐ C. Encourage students to represent their thinking while problem solving NOTE: Task must be a grade-level/developmentally-appropriate problem. That is, a solution is not readily apparent, the solution pathway is not obvious, and more than one pathway is possible. Comments:
2. Reason abstractly and quantitatively	☐ A. Facilitate opportunities for students to discuss representations or use representations to make sense of quantities and their relationships ☐ B. Encourage the flexible use of properties of operations, tools, and solution strategies when solving problems ☐ C1. Provide opportunities for students to decontextualize (abstract a situation) the mathematics within a mathematics task. ☐ C2. Provide opportunities for students to contextualize (identify referents for symbols involved) the mathematics within a mathematics task. NOTE: Must have C1 and C2 to receive credit for indicator. Comments:
3. Construct viable arguments and critique the reasoning of others	☐ A. Provide and orchestrate opportunities for students to listen to the solution strategies of others, discuss alternative strategies or solution(s), and defend their ideas ☐ B. Ask higher-order questions which encourage students to defend their ideas, consider student(s) response(s) before making code ☐ C. Provide prompts/tasks that encourage students to think critically about the mathematics they are learning, must be related to argumentation or proving events. ☐ D. Engage students in proving events that encourage students to develop and refine mathematical arguments (including conjectures) or proofs. Comments:
4. Model with mathematics	☐ A. Use mathematical models appropriate for the focus of the lesson ☐ B. Encourage student use of developmentally and content-appropriate mathematical models (e.g., variables, equations, coordinate grids) ☐ C. Remind students that a mathematical model used to represent a problem's solution is 'a work in progress,' and may be revised as needed ☐ D. Employ problems arising from everyday life, the local community, society, and workplace such that the solution is a model to reuse. NOTE: Must have D to be considered a task embedded within instruction promoting modeling with mathematics. Comments:
5. Use appropriate tools strategically	☐ A. Use appropriate physical and/or digital tools to represent, explore and deepen student understanding ☐ B. Help students make sound decisions concerning the use of specific tools appropriate for the grade level and content focus of the lesson ☐ C. Provide access to materials, models, tools, and/or technology-based resources that assist students in making conjectures necessary for solving problems. Students must use the resources. NOTE: Representations do NOT count as tools. Comments:
6. Attend to precision	☐ A. Emphasize the importance of precise communication by encouraging students to focus on clarity of the definitions, notation, and/or vocabulary used to convey their reasoning ☐ B. Encourage accuracy and efficiency in computation and problem-based solutions, expressing numerical answers, data and/or measurements with a degree of precision appropriate for the context of the problem ☐ C. Foster explanations and justifications using clearly articulated oral and/or written communication and grade-level appropriate conventions. Explanation or justification must go beyond IRE. Comments:
7. Look for and make use of structure	☐ A. Engage students in discussions emphasizing relationships between particular topics within a content domain or across content domains ☐ B. Recognize that the quantitative relationships modeled by operations and their properties remain important regardless of the operational focus of a lesson ☐ C. Provide activities in which students demonstrate their flexibility in representing mathematics in a number of ways e.g., 76 = (7 x 10) + 6; discussing types of quadrilaterals, etc. ☐ D. Encouraging examinations of a 'signal' and 'noise' in statistics-related tasks. Comments:
8. Look for express regularity in repeated reasoning	☐ A. Engage students in discussion related to repeated reasoning that may occur while executing a problem-solving strategy or in a problem's solution ☐ B. Draw attention to the prerequisite steps necessary to consider when solving a problem ☐ C. Urge students to continually evaluate the reasonableness of their results during problem solving Comments:

Figure 1. The *Revised SMPs Look-for Protocol*.

Outcomes

This section is named "outcomes" because it includes three parts stemming from the validation process. We share the qualitative results from stage six, then experiences coding data during stage seven, and finally, the quantitative findings from stage eight. The purpose for this is that the theme drawn from the expert panel and potential users came before quantitative analysis of the pilot data (i.e., stages six and eight, respectively). Our experiences learning as a team of coders, coding data, and coming together for interrater agreement have potential to inform the work of mathematics teacher educators. Thus, we aimed to remain faithful to the validation stages with respect to methodological fidelity.

Qualitative impressions from the expert panel and potential users: stage six

There was a single impression from inductive analysis of the expert panel and potential users. All members involved in the stages consistently agreed that the *Revised SMPs Look-for Protocol* provided a clear vision of gathering meaningful data about K–12 mathematics teachers' promotion of the SMPs. Many on the panel shared how the protocol offered a coherent set of observable aspects related to each SMP. A mathematics teacher who teaches grades 8 through pre-calculus and is a mentor teacher in her school commented, "This [revised protocol] is helpful for reflecting on what I could be doing in my classroom to promote the SMPs. I feel confident knowing that when I focus on one SMP that my principal, who is a former math teacher, could use this. In fact, I'd prefer that he use this over other observation tools required by our state because we could have a meaningful conversation about ways I might improve my instruction related to the math standards." This teacher's principal had previously served as a high-school mathematics teacher, reviewed the *Revised SMPs Look-for Protocol*, and added, "My teachers are expected to teach the standards. I want to be able to have a conversation with them about the ways they engage students in learning math found in the standards." The two mathematicians concurred that the *Revised SMPs Look-for Protocol* aligned with behaviors and habits typically seen by research-level mathematicians and was easy to use during their observations in K–12 school contexts. Relatedly, they shared how it helped them be more cognizant of mathematical behaviors they hoped to promote during their collegiate-level mathematics instruction. Finally, Fennell, a developer of the initial SMP look-for protocol and mathematics educator, shared that the additions found in the *Revised SMPs Look-for Protocol* allowed more teacher moves to be counted as promoting the SMPs, which did not hinder the quality of the observation or overall impressions of the teacher's instruction. Fennell expressed, "Allowing strategies and solutions to be counted as promoting SMP 1 is consistent with the literature on problems and problem solving. I'm glad it's there" (November 25, 2014). Drawing across these statements and others, we confidently drew the conclusion that the protocol addressed the SMPs and offered users a valid way to collect data about teachers' observable classroom behaviors related to promoting the SMPs.

Coding process: Stage seven

Our coding team consisted of 10 individuals, including two mathematics education faculty and eight graduate students. The story of learning to use the *Revised SMPs Look-for Protocol*, coming to acceptable interrater agreement, and coding videos is likely to inform other mathematics teacher educators involved in similar activities.

We begin by contextualizing the team's shared background. All of the coders had been mathematics teachers and spent at least one year and up to nine years in the K–12 classroom. They also either engaged in more than 40 hours of professional development on the SMPs and/or led professional development on the SMPs. Put simply, this coding team developed a fairly robust understanding of the SMPs, which aided in marking indicators appropriately and ultimately coming to acceptable agreement.

Next, we reflected collaboratively on potential intended meanings of the SMPs prior to any coding. There were numerous discussions about teachers' instructional behaviors during the teaching of mathematics and how these behaviors might suggest the presence of an indicator. Discussions often led members of the team to locate a video with such behaviors and review the video together as a coding team. It was through these discussions that our team came together around an idea. For instance, our team initially struggled with coding indicator 5c, "Provide access to materials, models, tools, and/or technology-based resources that assist students in making conjectures necessary for solving problems. Students must use the resources." Our discussions helped the group focus on instructional moments when a teacher provided access to manipulatives. Specifically, we found that teachers often made comments expressing access to materials and technology (e.g., calculators), but rarely did teachers do so to foster making conjectures to solve problems. Our team noted and discussed that many teachers told students to use manipulatives with the goal of learning to use it in a specific problem situation. Conversely, a few teachers gave students a set of manipulatives, encouraged them to use what they wanted, but what was different was that these teachers asked students to clarify why a tool was helpful for a given situation before using it in a problematic situation. These discussions about unique moments helped us understand better what to look for during live or video-recorded instruction.

Following numerous discussions, the coding team agreed to practice coding with videos from the various grade bands (e.g., K–3, 4–8, 9–12). The coding team began by watching a video-recorded middle-grades lesson. Each person coded the video independently, which included marking indicators and noting evidence of that indicator with a time stamp and brief statement or keyword. A code was applied when evidence was present that was suggestive of an indicator. Coders were instructed to mark an indicator as long as there was evidence of it. Put another way, no difference was noted between strong, moderate, or weak evidence of an indicator—all of these types of evidence had equal potential for an indicator. Next, the team collected their indicators into one form and coders compared their scores to the codes from the group (e.g., marking 5c or not marking 5c). We discussed areas where there was disagreement. At times, the video was watched again and arguments for or against an indicator were expressed until the group came to consensus. Then, one set of codes was agreed on. Individuals noted the agreed-on codes on their protocols. Finally, the coding team watched the video again with the agreed-on codes in front of them. This process was repeated for an early-childhood and high school video as well.

The final step of the seventh stage was coming to interrater agreement. Individuals were provided with a video, asked to code it independently from the other coders, and send a copy of their indicators to the first author within three days. As indicated earlier, our team arrived at 93% agreement. Afterwards, coders met informally to discuss ideas. The first author held a full-day meeting two more times with the coding team, because it was likely that coders might deviate during the two years of analysis. This meeting included discussing typical observations of instruction, commonly seen indicators, patterns across specific grade bands, and finally coding a video selected at random and discussing observed indicators. After the full-day meeting, coders watched a video independently and sent indicators back to the first author. Interrater agreement on the latter two occasions remained at or above 93%. In sum, the process of coding videos with the *Revised SMPs Look-for Protocol* required intense work by a team of individuals with a robust understanding of the SMPs.

Quantitative findings from reliability analysis: Stage eight

Multiple forms of reliability were explored as a means to gather sufficient psychometric evidence. The first form was internal consistency, which was acceptable with a Cronbach's alpha level of 0.801 for the overall measure. Cronbach's alpha levels between 0.70 and 0.90 are considered appropriate for assessments (Tavakol & Dennick, 2011). Measures with internal consistency below 0.70 could represent an assessment with poorly interrelated items, and a measure with internal consistency above 0.90 could possess too much item redundancy (Tavakol & Dennick, 2011).

The second form of reliability was test–retest reliability, which was acceptable with a correlation coefficient of 0.721 from pre-PD observation to post-PD observation. Test–retest reliability is considered to be one of the strongest forms of reliability evidence (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). This statistic suggests that teacher growth from pre- to post-PD is not always consistent across participants. Further investigation of the data clearly demonstrated this phenomenon, with those teachers performing higher at pre-PD demonstrating less growth at post-PD than teachers who performed lower over time. While conceptually it makes sense that teachers would have the ability to grow more if demonstrating lower levels of performance at pre-PD, it does not allow for production of what are considered good or excellent test–retest reliability coefficients.

Discussion

This study adds to the growing body of observation protocols validated for use in K–12 mathematics classrooms (see Lesseig et al., 2017; Boston et al., 2015 for a review) and builds on the Fennell et al. (2013) development of a tool to gather observational data about teachers' promotion of the SMPs. Results of our study fill a needed gap, as no validated tools currently focus on this area within K–12 instruction. Table 3 provides a comparison of the *Revised SMPs Look-for Protocol* and observation protocols mentioned in this article.

Mathematics teachers, curriculum leaders, and researchers may feel confident using this tool to explore the ways in which teachers foster the SMPs during instruction, and perhaps explore teachers' instructional changes using two or more observations. The present study provides evidence for validity and reliability for an observation protocol in K–12 mathematics classrooms specific to the SMPs, which are one-half of the content standards. It may behoove researchers and groups to examine protocols for validity evidence (e.g., Kane & Staiger, 2012) before allowing results from them to have serious consequences.

Mathematics teacher educators make choices in analytical tools, and there are a variety of lenses and available tools. Our study provides evidence for using the *Revised SMPs Look-for Protocol* as a tool in research and evaluative contexts of mathematics teacher education. Those aiming to affect teachers' practices through PD might consider using the lens of teachers' promotion of the SMPs. That is, use of the protocol could enable an investigator to gather data about changes in teachers' promotion of the SMPs before and after PD.

Table 3. Comparison of classroom observation protocols.

Classroom observation protocol	Domains measured	Medium	Training
Revised SMPs Look-for Protocol	Teacher facilitation.	Live observations. Videotape.	Not necessarily. Scorer must be knowledgeable about and have facility with the Standards for Mathematical Practice.
MCOP2	Student engagement. Teacher facilitation.	Live observations. Videotape.	Not necessarily. Scorer must be knowledgeable about and have facility with the Standards for Mathematical Practice.
RTOP	Lesson design and implementation. (Science and mathematics) content. Classroom culture.	Live observations. Videotape.	Yes. One-day video-based training prior to use. Individuals should have strong agreement or reliability with anchor lesson in training
IQA	Academic rigor. Rigor of mathematical discussions. Teacher's expectations.	Live observations. Videotape. Assignment collection.	Yes. Two-day training delivered by an IQA teach member.
MQI	Classwork is connected to mathematics. Richness of the mathematics. Working with students and mathematics. Errors and imprecision. Student participation in meaning-making and reasoning.	Videotape.	Yes. Online video-based training, which takes approximately 16 hours to complete.

Note. This table is modified from Sherman et al. (2014).

Future research

The results of this study offer validity evidence for the *Revised SMPs Look-for Protocol*, as part of the validity argument associated with this tool. We encourage mathematics education researchers to explore connections between the SMPs and the Mathematics Teaching Practices described in *Principles to Action: Ensuring Mathematical Success for All* (National Council of Teachers of Mathematics, 2014). It may be that teachers' promotion of the SMPs might be indicative of one or more Mathematics Teaching Practices, and this would imply a need for exploration of validity evidence related to other variables not considered in this study. Similarly, one might explore outcomes on the *Revised SMPs Look-for Protocol* and relationships to other variables, particularly teachers' content knowledge. We hypothesize that greater content knowledge for teaching may be correlated (convergent evidence) with higher scores on the *Revised SMPs Look-for Protocol*. A third area to explore is examining the degree to which the *Revised SMPs Look-for Protocol* and *MCOP*^{2} have concurrent validity with each other. That is, does use of one protocol lead to different results than use of the other? Boston et al. (2015) and Lesseig et al (2017) suggested such a review with the RTOP, IQA, and MQI, but results of this analysis have not been shared in this article. We conjecture that use of the *Revised SMPs Look-for Protocol* and *MCOP*^{2} should lead to the same result because they share a broadly framed construct of interest (i.e., SMPs observed during classroom instruction), but research is needed to respond to this inquiry.

Considerations

As a caveat, we do not advocate the use of the *Revised SMPs Look-for Protocol* by those unfamiliar with the SMPs or without a coherent understanding of each SMP. Hence a limitation of this protocol is that those without a deep understanding of the SMPs and K–12 mathematics instruction may not draw the same conclusions as others. A second consideration is that use of this protocol, as suggested in this study, is bounded by K–12 live and video-recorded mathematics instruction. We do not advocate using this protocol for observations in undergraduate and graduate-level mathematics and mathematics education courses without gathering sufficient validity evidence.

Final thoughts

The primary purpose of the study was to share information related to validating the *Revised SMPs Look-for Protocol* for the purpose of analyzing K–12 teachers' promotion of mathematical behaviors and habits framed by the CCSSM SMPs. We aimed to share validity evidence from cognitive interviews and the expert panel, our experiences coding and coming together as a research team interested in mathematics teacher education, as well as results from internal consistency and test–retest reliability analyses. A secondary aim was to educate the mathematics teacher education community about the potential uses of this tool for research and PD. In sum, a diverse audience may use the *Revised SMPs Look-for Protocol* to gather data about K–12 teachers' promotion of the SMPs during video-recorded or live classroom instruction. Relatedly, mathematics teacher educators might consider our experiences coding instruction under advisement as they use observation protocols in their research, program evaluation, and coaching of teachers.

Funding

This work was support by two generous Mathematics-Science Partnership grants from the Ohio Department of Education, #062839.

ORCID

Jonathan D. Bostic http://orcid.org/0000-0003-2506-0491

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.

Artino, Jr., A., La Rochelle, J., Dezee, K., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE guide no 87. *Medical Teacher*, *36*, 463–474. doi:10.3109/0142159X.2014.889814

Bostic, J. (2015). A blizzard of a value. *Mathematics Teaching in the Middle School*, *20*, 350– 357.

Bostic, J., & Matney, G. (2014). Role-playing the Standards for Mathematical Practice: A Professional Development Tool. *Journal for Mathematics Education Leadership*, *15*(2), 3–10.

Bostic, J., & Matney, G. (2016). Leveraging modeling with mathematics-focused instruction to promote other standards for mathematical practice. *Journal of Mathematics Education Leadership*, *17*(2), 21–33.

Boston, M., Bostic, J., Lesseig, K., Sherman, M. (2015). Classroom Observation tools to support the work of mathematics teacher educators. *Invited manuscript for Mathematics Teacher Educator*, *3*, 154–175.

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.

Boston, M. D., & Wolf, M. K. (2006). *Assessing academic rigor in mathematics instruction: The development of instructional quality assessment toolkit* (CSE Tech. Rep. No. 672). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Common Core State Standards Initiative. (2010). *Common core standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and Council of Chief State School Officers.

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory* (2nd ed.). Mason, OH: Wadsworth.

Fennell, F., Kobett, E., & Wray, J. (2013, February). *Using look fors to consider the Common Core content standards*. Presentation at the Annual Meeting of the Association of Mathematics Teacher Educators, Orlando, FL.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education report: A preK-12 framework*. Alexandria, VA: American Statistical Association.

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

Glaser, B., & Strauss, A. (1967/2012). *The discovery of grounded theory: Strategies for qualitative research*. Mill Valley, CA: Sociology Press.

Gleason, J., & Cofer, L. D. (2014). Mathematics classroom observation protocol for practices results in undergraduate mathematics classrooms. In T. Fukawa-Connelly, G. Karakok, K. Keene, & M. Zandieh (Eds.), *Proceedings of the 17th Annual Conference on Research on Undergraduate Mathematics Education* (pp. 93–103). Denver, CO.

Gleason, J., Livers, S., & Zelkowski, J. (2017) Mathematics Classroom Observation Protocol for Practices (MCOP2): A validation study, *Investigations in Mathematics Learning*, *9*(3), 111–129.

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

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, *26*, 430–511. doi:10.1080/07370000802177235

James, L., Demaree, R., & Wolf, G. (1993). R_{wg}: An assessment of within-group interrater agreement. *Journal of Applied Psychology*, *78*, 306–309. doi:10.1037/0021-9010.78.2.306

Kane, T., & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* (MET Project research paper, Bill and Melinda Gates Foundation). Retrieved from http://files.eric.ed.gov/fulltext/ED540962.pdf, Retrieved May 2016.

Kanold, T., & Larson, M. (2012). *Common Core mathematics in a PLC at work: Leader's guide*. Bloomington, IN: Solution Tree Press.

Koestler, C., Felton, M. D., Bieda, K. N., & Otten, S. (2013). *Connecting the NCTM Process Standards and the CCSSM Practices*. Reston, VA: National Council of Teachers of Mathematics.

Lesseig, K., Bostic, J., Sherman, M., & Boston, M. (2017, April). Classroom observation protocols: Choose your own tool. Paper presented at the meeting of the National Council of Teachers of Mathematics research conference. San Antonio, TX.

Matsumura, L. C., Garnier, H., Slater, S. C., & Boston, M. D. (2008). Toward measuring instructional interactions "at-scale." *Educational Assessment*, *13*, 267–300. doi:10.1080/10627190802602541

National Council of Teachers of Mathematics. (2014). *Principles to action: Ensuring mathematical success for all.* Reston, VA: Author.

Ohio Department of Education. (2013). *Typology of school districts*. Retrieved from http://education.ohio.gov/Topics/Data/Frequently-Requested-Data/Typology-of-Ohio-School-Districts, retrieved June 2017.

Sawada, D., Piburn, M., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2000). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics Journal*, *102*, 245–253. doi:10.1111/j.1949-8594.2002.tb17883.x

Smith, M., Jones, F., Gilbert, S., & Wieman, C. (2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE-Life Sciences Education*, *12*, 618–627. doi:10.1187/cbe.13-08-0154

Sherman, M., Bostic, J., Lesseig, K., & Boston, M. (2014, February). A comparison of commonly used mathematics classroom observation protocols. Paper presented at the meeting of the Association of Mathematics Teacher Educators. Irvine, CA.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. doi:10.5116/ijme.4dfb.8dfd

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.