EDITORIAL

3  Who Wants to Be an MTE Editor? A Goal for Your Professional Bucket List
Sandra Crespo

ARTICLES

6  Making the Most of Teacher Self-Captured Video
Elizabeth A. van Es, Shari L. Stockero, Miriam G. Sherin, Laura R. Van Zoest, and Elizabeth Dyer

20  Mentor-Guided Lesson Study as a Tool to Support Learning in Field Experiences
Kristen N. Bieda, Jillian Cavanna, and Xueying Ji

32  Transforming Perceptions of Proof: A Four-Part Instructional Sequence
Justin D. Boyle, Sarah K. Bleiler, Sean P. Yee, and Yi-Yin (Winnie) Ko

71  Enhancing Teachers’ Assessment of Mathematical Processes Through Test Analysis in University Courses
Patricia D. Hunsader, Barbara Zorin, and Denisse R. Thompson

93  Developing a Mathematics Instructional Practice Survey: Considerations and Evidence
Michele B. Carney, Jonathan L. Brendselur, Gwyneth R. Hughes, and Keith Thiede

119  CAEP Program Reviewer Recognition
Address all correspondence to Mathematics Teacher Educator, 1906 Association Drive, Reston, VA 20191-1502. Prepare manuscripts according to the Publication Manual of the American Psychological Association. No author identification should appear on the manuscript; the journal uses a blind-review process. To submit a manuscript, access mte.msubmit.net.

Mathematics Teacher Educator (MTE) (ISSN 2167-9789) is a joint venture between the National Council of Teachers of Mathematics and the Association of Mathematics Teacher Educators. MTE is published online only in September and in March by the National Council of Teachers of Mathematics, 1906 Association Drive, Reston, VA 20191-1502. © 2015. The National Council of Teachers of Mathematics.

Mission and Goals of MTE: The Mathematics Teacher Educator will contribute to building a professional knowledge base for mathematics teacher educators that stems from, develops, and strengthens practitioner knowledge. The journal will provide a means for practitioner knowledge related to the preparation and support of teachers of mathematics to be not only public, shared, and stored, but also verified and improved over time (Hiebert, Gallimore, & Stigler 2002).

NCTM BOARD OF DIRECTORS
Diane J. Briars, Pittsburgh, Pennsylvania; President
Matthew R. Larson, Lincoln Public Schools, Nebraska; President-Elect
Robert M. Doucette, NCTM, Executive Director
Jennifer Bay-Williams, University of Louisville, Kentucky
Nadine Bezuk, San Diego State University, California
Florence Glanfield, University of Alberta, Edmonton
Paul Kelley, Anoka High School, Anoka, Minnesota
Cathy Martin, Denver Public Schools, Colorado
Ruth Harbin Miles, Falmouth Elementary School, Stafford, Virginia
Jane Porath, Traverse City East Middle School, Michigan
John SanGiovanni, Howard County Public Schools, Maryland
Denise A. Spangler, University of Georgia, Athens
Marilyn E. Strutchens, Auburn University, Alabama
Trena L. Wilkerson, Baylor University, Waco, Texas
Rose Mary Zbiek, Pennsylvania State University, University Park

AMTE BOARD OF DIRECTORS
Christine Thomas, Georgia State University, Atlanta; President
Fran Arbaugh, Pennsylvania State University, University Park; Past President
Tim Hendrix, Meredith College, Raleigh, North Carolina; Executive Director
Babette Benken, California State University, Long Beach
Tim Boerst, University of Michigan, Ann Arbor
Christine A. Browning, Western Michigan University, Kalamazoo

MTE EDITORIAL PANEL
Sandra Crespo, Michigan State University, East Lansing; Editor
Kristen Bieda, Michigan State University, East Lansing; Associate Editor
Christine A. Browning, Western Michigan University, Kalamazoo, AMTE Board Liaison
Rebekah Elliott, Oregon State University, Corvallis
Anthony M. A. Fernandes, University of North Carolina, Charlotte
Randall Groth, Salisbury University, Maryland
Amy Hillen, Kennesaw State University, Georgia
Jeffrey C. Shih, University of Nevada, Las Vegas
Laura R. Van Zoest, Western Michigan University, Kalamazoo; Chair

AMTE BOARD OF DIRECTORS
Christine Thomas, Georgia State University, Atlanta; President
Fran Arbaugh, Pennsylvania State University, University Park; Past President
Tim Hendrix, Meredith College, Raleigh, North Carolina; Executive Director
Babette Benken, California State University, Long Beach
Tim Boerst, University of Michigan, Ann Arbor
Christine A. Browning, Western Michigan University, Kalamazoo

Megan Burton, Auburn University, Alabama
Joe Champion, Boise State University, Idaho
Susan Gay, University of Kansas, Lawrence
Suzanne R. Harper, Miami University, Oxford, Ohio
Kathleen Lynch-Davis, Appalachian State University, Boone, North Carolina
Nicole Rigelman, Portland State University, Oregon
Edward A. Silver, University of Michigan, Ann Arbor
Dorothy Y. White, University of Georgia, Athens

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS

Mathematics Teacher Educator • Vol. 4, No. 1, September 2015
Who Wants to Be an MTE Editor? A Goal for Your Professional Bucket List

Sandra Crespo
Editor, Mathematics Teacher Educator

Imagine: An email pops up in your inbox inviting you to apply to be MTE’s next editor. How would you react? Would you jump at the opportunity? Need to think more about it? Respond with a definite “no”? What information would you need in order to help you decide? In my first editorial, I share a bit of why I chose to jump at the opportunity to become MTE editor and what I have learned in the last year as Editor Designate, processing manuscripts alongside the founding and former MTE editor Peg Smith. In addition to providing insight into the MTE review process, I hope this editorial convinces many of you to seriously consider jumping on the opportunity to become MTE’s next editor when it comes around again in 3 years.

As readers, authors, and reviewers of manuscripts without editing experience, we tend to have a limited perspective of the peer review process and the role of editor. The image we might have of a journal editor as a grumpy judge who decides the fate of manuscripts and finds that nothing is ever good enough, does not make the role of editor seem very appealing. Although there may be editors who operate in this fashion, this is largely a distortion of what a journal editor is and does. A different, perhaps more accurate representation of a journal editor is that of a coach with a passion for mentoring players and for making the game worth playing. This image of a journal editor better fits the mission of MTE as a journal that seeks to improve the knowledge base and the practices of mathematics teacher educators.

When considering the invitation to apply for the position of MTE editor, I did not just jump in with eyes closed. I first made an appointment with editor Peg Smith to learn more about the journal and its mission. In that first conversation, I learned pretty quickly that the mission of the journal is to serve not only the readers of the journal but also the authors and reviewers who contribute to the journal, as they too are crucial to the journal’s mission to build on and improve the practice of mathematics teacher educators. It became much clearer that being MTE editor shared many similarities with my other work as a mathematics teacher educator working with multiple constituencies and communities. Once I made that connection, the decision to apply for the MTE editorship made it to the top of my professional bucket list.

Another reason that becoming MTE’s next editor was attractive to me was doing with my own experience as a scholar in mathematics education. As an author of journal manuscripts, I have received a full spectrum of review decisions—reject, revise and resubmit, and accept. I have experienced firsthand the joys and tribulations that go with writing for publication and the key role that journal editors have played in my development as a writer of scholarly articles for different types of research and practitioner journals. I have benefited from the generosity of strangers who have reviewed my work and invested their time and energy in giving advice on how to sharpen and strengthen the coherence of the arguments I was trying to make. More importantly, these scholars influenced my perspective on and approach to doing high-quality scholarly work in mathematics teacher education. Becoming the next MTE editor was my opportunity to pay forward the benefits I had received as a junior scholar in mathematics education.

Having been at the receiving end of editors’ decision letters, I came into this new role with some appreciation for how skillful journal editors recast, revoice, and summarize the key issues reviewers have identified that merit revision or rejection of a manuscript. How editors write manuscript decision letters to authors, especially first-time authors, matters not only because they determine the fate of the manuscript under review, but because their approach shapes how these scholars will then engage in the review process as authors and as reviewers of others’ manuscripts. Having spent time this year writing decision letters alongside Peg Smith and Melissa Boston (who served as associate editor), I have an even greater appreciation for this work, and I can understand why so many authors have written thank-you letters to them. They have set a high standard and model for what an editor’s letter looks like and demonstrated a commitment to doing this consistently over time. Both Kristen Bieda (who is now serving as associate editor) and I are committed to continuing their legacy and growing our practice as editors who will write substantive and constructive decision letters to authors.

The opportunity to spend a year working closely with Peg Smith and the journal’s Editorial Panel has supported my successful transition to the role of editor and has given...
me many important insights about journal editing and the editorial decision-making process. I share here three important realizations I have had related to the role of MTE editor and the exciting work I have signed up to do for the next 3 years. The first one is that editing a journal requires collaboration, the second is the importance of a transparent and educative review process, and the third is that MTE is an important resource for current and future generations of mathematics teacher educators.

Journal editing is a collaborative practice.
Regardless of which image one holds for a journal editor, either as judge or coach, both require interactions with multiple collaborators—the author, reviewers, the associate editor, guest editors, and review board members. The social and communicative interactions and skills that are required to effectively manage these relationships resemble the demands of doing productive collaborative work—no one is allowed to take over, and everyone’s work is important to the success of the group’s product. More important, it is not possible for a single person to do the work of journal editing alone. Editing a journal requires the smarts and the skills of multiple people. The editor cannot do everything that needs to happen for a quality publication to come to fruition. Notice that most journals, MTE included, have an associate editor. Kristen Bieda will be co-editing the MTE journal with me and has been managing manuscripts since we began our term this May.

Everyone involved in producing a journal needs to be invested in the quality of the product because it represents all of our vision and our hard work, not just that of the editor’s. Even this editorial is undergoing a collaborative review and revision process; I asked MTE associate editor Kristen Bieda and Editorial Panel Chair Laura Van Zeeost to review it, and they provided me with critical and constructive feedback that I used to revise and improve it.

The peer-review process should be transparent and educative.
Another important goal of MTE and its editors is to make the peer review process transparent and educative. Making the MTE review process transparent means that there should be no mystery about what happens to manuscripts once they are submitted for review. Of course, the peer review is blinded and reviewers’ names are not disclosed to authors, but the process of assigning reviewers to manuscripts, the reviewers’ comments, and the editor’s decision about the submission should all be accessible to the authors.

MTE works to make the process transparent to prospective authors not only by providing information for authors online but also by actively sharing information with them through webinars on how to write for MTE and sessions at the Association of Mathematics Teacher Educators (AMTE) and at the NCTM Annual Meeting and Exposition focused on writing manuscripts for MTE. This past year at AMTE, we offered a session featuring published MTE authors who successfully turned their AMTE conference presentation into an MTE publication. At the NCTM Research Conference we featured authors who had successfully taken ideas from their research work and crafted them into practitioner manuscripts that were then published in MTE.

MTE also strives to provide educative reviews for all authors, even those whose manuscripts are rejected. This requires developing reviewers’ capacity to produce good reviews. Skilled reviewers and high-quality reviews are crucial to the peer-review process. Reviewers who meet deadlines and are clear, thoughtful, and sensitive in their comments to the author are worth their weight in gold. Good reviewers do not try to rewrite the manuscript for the author, but instead provide feedback that will bolster the author’s ideas and arguments. They identify the parts of the manuscript that meet the review criteria for the journal and those where the manuscript falls short or can be strengthened. These reviews are not only helpful to the editor and to the author of the manuscript, they are also educative to other reviewers of the same manuscript.

MTE editors are committed to providing all authors with educative reviews, including those whose manuscripts have been rejected. Each manuscript is guaranteed to have at least one very strong review because one editorial panel member is always assigned as a reviewer. Editorial board members discuss and clarify the journal’s review criteria and how to write helpful reviews by reviewing manuscripts together at the annual board meeting and when new members are appointed to the board. This distributes and builds the reviewing skills of everyone on the panel. Other reviewers, especially those who review frequently, also benefit from reading the editor’s decision letters and the panel members’ reviews.

Authors whose manuscripts are returned without review because they do not fit the journal’s criteria are also provided with an explanation of why their submission does not fit the journal and with suggestions on where they could submit their manuscript. Manuscripts that clearly fit the journal but are missing criteria or are underdeveloped in several of the review criteria are quickly returned to the authors for further development. They are provided with an editor’s decision letter that clearly identifies the weaknesses of the manuscript relative to the MTE review criteria and with specific suggestions for improving their manuscript either to submit to another journal or back to MTE as a new manuscript submission.
MTE strives to support the next generation of mathematics teacher educators.

MTE is committed to developing the next generation of mathematics teacher educators by publishing articles that address shared problems of practice and that allow current and future generations to replicate and build on the solutions shared by the authors. MTE editors make visible through their editorials the relevance and potential impact of the published articles for current and future educators. In this issue, which inaugurates my and Kristen’s tenure as the editors of MTE, we can see evidence of this commitment. Although the previous editorial team handled all five articles in this issue, the same commitment will continue under the next editorial team’s leadership. Each article in this issue provide readers with theoretical and practical tools to address current issues in the work of mathematics teacher educators, but because they target vexing problems of professional practice, these articles will continue to have relevance for future generations of mathematics teacher educators reading the archived issues of this journal.

In “Making the Most of Teacher Self-Captured Video,” van Es and colleagues discuss important questions and strategies for using teacher self-captured videos in professional development settings. In “Mentor-Guided Lesson Study as a Tool to Support Learning in Field Experiences,” Bieda and colleagues share a creative approach to engage mentor teachers and preservice teachers in a field-related course assignment. In “Transforming Perceptions of Proof: A Four-Part Instructional Sequence,” Boyle and colleagues describe an instructional sequence that broadens and deepens teachers’ perception of the nature of proof. In “Enhancing Teachers’ Assessment of Mathematical Processes Through Test Analysis in University Courses,” Hunsader and colleagues detail how to enhance preservice and in-service teachers’ knowledge of classroom assessments. And in “Developing a Mathematics Instructional Practice Survey: Considerations and Evidence,” Carney and colleagues give us insights into how to design and test survey tools that can be used to seriously explore instructional practice at scale.

I am thrilled to be writing my first editorial with this great collection of articles. I am very honored to have been entrusted to guide and steer MTE’s journey into its next 3 years and to be working alongside Kristen as associate editor, with a very committed and talented group of editorial board members. Kristen and I are looking forward to continuing to build the journal’s visibility and reputation as a publication venue that showcases the good work that mathematics teacher educators are doing and that will be admired and built upon by the next generation of mathematics teacher educators. More important, we are committed to making the editorship of MTE an attractive professional bucket list item for scholars for many generations to come.

To close, I hope this editorial has encouraged many of you to add MTE editorship to your professional bucket lists. If so, you might want to consider what will help you prepare for that role and what opportunities you should be actively seeking that will stretch your reviewing, mentoring, and leadership skills. You might start by learning more about the writing process or about how to provide mindful feedback to authors. Working on these skills will not only prepare you for a future role as editor of this journal but will also improve your own scholarly writing. Serving on editorial boards of peer-reviewed journals and books is another way to prepare for the role of future MTE editor and get an insider’s look into journal editing. Kristen and I look forward to mentoring and passing on what we have learned to the next editorial team. In the meantime, we are very excited to be the new editors of MTE and to contribute our perspectives to the journal’s mission “to build a professional knowledge base for mathematics teacher educators that stems from, develops, and strengthens practitioner knowledge.”

Author
Sandra Crespo, Department of Teacher Education, Michigan State University, East Lansing, MI 48823; crespo@msu.edu
Making the Most of Teacher Self-Captured Video*

Elizabeth A. van Es  
*University of California, Irvine*

Shari L. Stockero  
*Michigan Technological University*

Miriam G. Sherin  
*Northwestern University*

Laura R. Van Zoest  
*Western Michigan University*

Elizabeth Dyer  
*Northwestern University*

Recent advances in technology have resulted in an array of new digital tools for capturing classroom video, making it much easier for teachers to collect video from their own classrooms and share it with colleagues, both near and far. We view teacher self-captured video as a promising tool for improving mathematics teacher education. In this article, we discuss three issues that are essential for making the most of self-captured video: camera position, how much video to capture, and when to specify tasks for capturing, selecting, and using video. We propose that the act of deliberately participating in the self-capture process, as well as viewing and analyzing one’s own video with colleagues, offers worthwhile opportunities for mathematics teacher learning.

**Key words:** Teacher professional development; Self-capture video; Video analysis

Video has been used extensively in mathematics teacher education for more than 2 decades, and its use continues to be widespread today. Video captures much of the complexity of teaching and allows teachers an opportunity to view and examine instruction with the guidance of mentors and colleagues. In the past, video-based professional development typically relied on researchers or teacher educators to videotape in teachers’ classrooms. Such work led to the development of a range of published curriculum materials for mathematics teachers (e.g., Boaler & Humphreys, 2005; Goldsmith & Seago, 2012; Seago et al., in press; Seago, Mumme, & Branca, 2004), as well as programs in which researchers selected video from participants’ own classrooms for participants to view with colleagues (e.g., Borko, Jacobs, Eiteljorg, & Pittman, 2008; Grant, Kline, & Van Zoest, 2001; Sherin & Han, 2004; van Es & Sherin, 2008). These approaches have provided teachers with valuable opportunities to examine student mathematical thinking, new visions of mathematics teaching, and classroom discourse.

Recent advances in technology support a new approach to video-based professional development in which video from teachers’ classrooms is captured by the teachers themselves. Because teachers now have access to small, high-quality consumer cameras (including phones and tablets), they no longer need specialized equipment for videotaping in their classrooms. In addition, new online applications make it relatively easy to upload and share video with colleagues. So what does this mean for mathematics teacher education? What might self-captured video have to offer teachers and teacher educators?

First, self-captured video increases the ease with which teachers can engage in ongoing cycles of planning, teaching, and reflection as recommended by current research initiatives (Hiebert, Morris, & Glass, 2003; Koellner et al., 2007; McDonald, Kazemi, & Kavanagh, 2013). Second, having teachers capture video from their own classrooms reduces the burden on teacher educators, making video-based programs more feasible to implement at scale. Third, being able to share video of one’s teaching with peers quickly and directly increases the potential for teacher involvement in new forms of professional learning communities, both local and at a distance (see Krammer et al., 2006; Lieberman & Pointer-Mace, 2009). Furthermore, we conjecture that the act of capturing video provides teachers with learning opportunities beyond those of reflecting on and analyzing video from one’s classroom.

Given recent research and policy that advocate for mathematics teachers to attend closely to student thinking and engage in practices that are responsive to student ideas (e.g., Kazemi, Franke, & Lampert, 2009; National Council of Teachers of Mathematics, 2014; Sherin, Jacobs, & Philipp, 2011), we believe that teacher self-captured video is a particularly worthwhile tool for mathematics teacher education and professional development. Teachers can

---

*In accordance with MTE policy regarding conflicts of interest with the editor, the review process for this manuscript was handled by Denisse R. Thompson, University of South Florida. This article was submitted and accepted under the editorship of Margaret Smith.*
deliberately plan to both elicit student thinking during a lesson and capture those moments on video. Research finds that such purposeful planning for video capture, as well as actively capturing moments during instruction, can fundamentally shift the opportunities that teachers create for student mathematical thinking to emerge and the ways in which teachers take notice of these ideas during instruction (Sherin, Russ, & Colestock, 2011; Sherin, Russ, Sherin, & Colestock, 2008). Furthermore, although video of others (e.g., Boaler & Humphreys, 2005; Seago et al., 2004) has been used successfully to promote teacher attention to and interpretation of student thinking (Goldsmith & Seago, 2011), research has found value in mathematics teachers examining their own students’ thinking through videos from their own classrooms (van Es & Sherin, 2010). Self-captured video extends this perspective by promoting teacher attention to student thinking during both the capture and the reflection periods (Dyer, 2013).

Although there is clearly great potential in using self-captured video for teacher learning, realizing this promise requires consideration of a variety of issues. Mathematics teacher educators need to give careful thought to how and why self-captured video is used: How can teacher educators help teachers collect video that is usable and substantive? What kinds of tasks are productive for teachers to engage in around self-captured video and when should these tasks be specified? What methods should be used to organize sharing and discussion of video? To what extent will professional development around the video be facilitated and by whom? In this article, we begin to address questions about self-captured video by examining the following three issues, each of which we believe is essential to making the most of self-captured video: (a) where to position equipment for optimal results, (b) how much video to capture and share with colleagues, and (c) when to specify tasks for using teacher-captured video.

**Our Background With Teacher Self-Captured Video**

Our experience using video with teachers is extensive, working with both preservice and in-service teachers and spanning more than 20 years. In terms of teacher self-captured video, our experience is quite diverse. Across the five authors of this article, we have used a range of camera types, including stationary, hand-held, and wearable cameras. In addition, we have asked teachers to capture different types of video, from whole lessons to shorter segments, and have provided varied degrees of structure and guidance for video capture and analysis tasks. Across these experiences is a shared interest in using teacher self-captured video to help teachers consider student mathematical thinking as it relates to the practice of teaching (see for example Dyer, 2013; Stockero, 2013, 2014) and to develop teachers’ agency in their own professional development. For example, several of us work with preservice mathematics teachers, helping them to develop video in their field placement assignments to develop their attention to and interpretation of student thinking to inform planning and in-the-moment teacher actions. We have also worked with in-service mathematics teachers who capture videos of their own classrooms for various video-based programs, including National Board Certification. In all these contexts, teachers typically engage in collaborative analysis or discussions of their videos, facilitated by themselves, a teacher leader, or an outside facilitator. Throughout this article, we draw on data from our research with these teachers—including observations of professional development and teacher education courses, interviews with teachers, and teachers’ written work—to illustrate and contextualize our claims.

As previously stated, our goal is to share some of our insights concerning the decisions that teacher educators need to make when planning to have mathematics teachers capture video of their own instruction. Regardless of whether the use of self-captured video is highly structured by an external facilitator or informally teacher-led, we think these are important issues to consider. In particular, we discuss the affordances and constraints of different choices related to camera positioning, how much video to capture, and when to specify tasks for capturing, selecting, and using the video. We recognize that issues related to equipment selection are also important. Although equipment selection is not a central focus of this article, we include details about various types of cameras and accessories, such as microphones and tripods, in Appendix A.

**Three Issues in Teacher Self-Captured Video**

**Positioning of Equipment**

When teachers videotape themselves, it is nearly impossible to obtain access to every element of the classroom environment. Instead teachers need to make choices about particular classroom features on which to focus. Although video is often believed to be objective—capturing activities from a neutral perspective—different positioning of equipment can cause various aspects of the same classroom lesson to appear more or less salient (Goldman, 2007; Roth, 2007). Therefore, how video equipment is positioned has important consequences for what can be seen in the video. Of particular interest to us are the ways in which video can provide access to the teacher...
and the students in the classroom and, as a result, to the mathematics the class is exploring.

**Access to teachers.** In our work with both preservice and in-service mathematics teachers, we find that teachers often place the video camera at the back of the room when first videotaping. Their reasons for doing so seem related to the fact that teachers often associate videotaping with evaluating or analyzing the teacher. As one teacher explained,

> When I think of colleagues taping themselves and sitting and discussing with each other, the first thing that I thought of was [that] we are going to look at me as a teacher and talk about what I'm doing as a teacher, not about what students are thinking.

For these teachers, placing the camera at the back of the room can be seen as a strategic choice because that position provides the best access to the teacher throughout the lesson. Other positions are also possible, however, as shown in Table 1. For example, a stationary camera positioned at the front of the room typically provides better access to the teacher’s interactions with individual or small groups of students, as long as the teacher is not standing too close to the front of the room. However, a camera positioned at the front of the room often will not capture the teacher when she is writing on the board or using a document camera or overhead projector. With regard to audio, positioning the camera at either the front or the back of the room can also make it difficult to hear the teacher’s interactions with individual or small groups of students over the noise of the classroom.

Sometimes it may be useful to position a camera on a small group of students rather than on the class as a whole. A constraint of this position is that it provides access to the teacher only when the teacher approaches that group. An affordance is that it allows the viewer to make sense of the teacher-student interactions in light of conversations among the students before and after the teacher approached the group. As one teacher wrote, “[It] worked really well to tape the group. I liked that I could see what they did after I left.”

Finally, wearable cameras provide no visual access to the teacher—the teacher is literally wearing the camera and is thus not seen in the video at all. Yet video taken from this perspective does provide a different kind of access to the teacher—access to what the teacher sees throughout the course of a lesson, as well as what a teacher hears as instruction takes place. One teacher commented, “It’s like [a] whole new way to see myself teaching.” In addition, a wearable camera provides excellent access to what the teacher says because the camera is so close to the teacher. To obtain similar auditory access to the teacher with other camera positions, some teachers choose to wear a wireless lapel microphone.

**Table 1**

*Equipment Position and Access to Teachers*

<table>
<thead>
<tr>
<th>Camera or microphone position</th>
<th>Visual affordance</th>
<th>Visual limitation</th>
<th>Auditory affordance</th>
<th>Auditory limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-of-the-class stationary</td>
<td>Teacher visible throughout classroom</td>
<td>Limited access to teacher interactions with individual or small groups of students</td>
<td>Teacher usually audible</td>
<td>Teacher interactions with students at the front of the room may be difficult to hear</td>
</tr>
<tr>
<td>Front-of-the-class stationary</td>
<td>Teacher visible when working with individual or small groups of students</td>
<td>Limited access to teacher at the front of the room</td>
<td>Teacher usually audible</td>
<td>Teacher interactions with students at the back of the room may be difficult to hear</td>
</tr>
<tr>
<td>Small group stationary</td>
<td>Teacher interactions with small groups of students visible</td>
<td>Limited access to teacher</td>
<td>Teacher usually audible when speaking to small groups</td>
<td>Teacher interactions with other groups may be difficult to hear</td>
</tr>
<tr>
<td>Teacher wearable</td>
<td>Camera reveals what teacher sees and hears</td>
<td>Teacher not visible</td>
<td>Teacher always audible; access to what teacher hears</td>
<td>Access only to students whom the teacher hears</td>
</tr>
</tbody>
</table>
Access to students. Over time, we find that teachers generally become more open to various camera placements, particularly when the goal is to see and hear students (see Table 2). Teachers quickly discover that positioning a camera in the back of the classroom has numerous limitations in gaining access to students: “I was trying [to look] at the students, but I couldn’t tell who was who. I heard what they said, but I couldn’t match it [with their faces].” With a camera placed at the back of the room, it is difficult to recognize if students are engaged, to see their gestures, or to see their reactions to events taking place. As with cameras in the back of the room, the front-of-the-room position provides a general sense of what is taking place in the classroom, but a camera in the front provides better access to students’ faces. A constraint of the front-of-room position is that it can be difficult to see work being done at the front of the classroom; teachers have been able to compensate for this by taking photographs of the board or using an interactive white board that can automatically save electronic copies of what is written on it.

In terms of audio, a microphone located at the front of the room, rather than at the back, provides much better access to what students say during whole-class formats because students often speak toward the front of the room. In both positions, however, the noise of the class as a whole can make it difficult to hear what students say as they work in small groups. In general, we have found that external microphones provide better quality audio and greater flexibility in microphone placement, particularly when placed near the students the teacher wishes to hear. A teacher explained, “I think the external microphone helped a lot with the audio. I never had any problems hearing the students talking, and my class got pretty loud at times. I also liked being able to put the camera at the back of the classroom and put the microphone on a table in the middle of the classroom to record whole class discussions.” In general, we have found that the teacher typically can be heard from most positions, while students’ voices are often difficult to hear, particularly when the microphone is placed at the back of the room.

A stationary camera positioned directly on a group of students offers a different perspective. In this setup, the viewer has extensive access to those students who appear in the camera’s field of view, particularly in terms of facial expressions and gestures. The rest of the students in the class are not visible, however. When focusing on a small group, placing an external microphone in the center of a group can help capture what students say because the microphone on a camera is often insufficient for capturing individual voices over the general noise of the classroom during small group discussions. Moreover, when capturing video of a small group, it can be vital to have a copy of the students’ work to make sense of what they are doing because students often reference their written work when sharing ideas with their classmates.

Wearable teacher cameras provide different access to students. When the teacher is interacting with the class as a whole and scanning the room, students throughout the classroom will likely be captured, and it will be students’ faces that are visible. Teachers like to remind us that “[With the wearable] camera you have to look at students [in the video]. That’s all that’s there.” In addition, when the teacher approaches an individual student, or a small group of students, a more intimate view of those students is offered, in line with how closely the teacher and students are interacting. In fact, with high-resolution cameras it is often

<table>
<thead>
<tr>
<th>Camera or microphone position</th>
<th>Visual affordance</th>
<th>Visual limitation</th>
<th>Auditory affordance</th>
<th>Auditory limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-of-the-class stationary stationery</td>
<td>Many students appear in video</td>
<td>Students’ faces may not be visible</td>
<td>Hear students in back of room</td>
<td>Students in front of room often hard to hear</td>
</tr>
<tr>
<td>Front-of-the-class stationary</td>
<td>Many students’ faces are visible</td>
<td>May not see students if they present at front of the room</td>
<td>Students usually audible during whole-class format</td>
<td>Students at back of room often hard to hear</td>
</tr>
<tr>
<td>Small group stationary</td>
<td>Sustained access to group of students</td>
<td>Few students appear in video</td>
<td>Easily hear students in the group</td>
<td>Hard to hear students not in the group</td>
</tr>
<tr>
<td>Teacher wearable</td>
<td>Students central focus of video</td>
<td>Access only to those students in teacher’s field of view</td>
<td>Easily hear students close to teacher</td>
<td>Access only to students whom the teacher hears</td>
</tr>
</tbody>
</table>
Making the Most of Teacher Self-Captured Video

possible to see the student written work that the teacher is examining. Finally, with the location of the microphone on the teacher, the sound quality is good, particularly for those students who are close to the teacher.

Determining the Amount of Video to Capture

A second issue that has become relevant in our work is deciding how much video to capture. This issue is important because it has implications for what teachers have access to view and analyze, as well as for when in the video capture process they have to devote the most time, energy, and careful thought. Teachers might capture an entire lesson, define selected excerpts to capture prior to a lesson, or, if using a wearable camera, capture events and interactions in an impromptu fashion during a lesson. We discuss the affordances and limitations of each approach, particularly as they relate to sharing and analyzing video with others.

Whole lesson. In our work with teachers, we find that capturing a whole lesson offers several affordances (see Table 3). First, it decreases the decisions teachers need to make about what to capture both before and during their teaching of the lesson being captured. Preservice teachers, in particular, are often challenged by selecting, designing, and organizing materials for use in the lesson and are also overwhelmed by all the decision points as they teach their lesson. Being able to set up a camera and press the record button at the start and end of a lesson requires less cognitive demand and allows teachers to focus on their teaching. Another affordance of capturing a whole lesson is that the context of the entire lesson is available. Teachers can see how a lesson begins, builds, transitions, and concludes, and can examine how these pieces fit together to support or constrain student learning. A third affordance is that colleagues can view an entire lesson from one another’s classrooms. This is particularly worthwhile when teachers plan a full lesson together, as in a lesson study approach (Lewis, Perry, & Hurd, 2009). If they are unable to be present to watch the lesson in real time, they can later view their colleagues teaching the lesson and then collectively discuss strengths of the lesson and areas for improvement. Finally, with a whole lesson, teachers can decide what to share after they have an opportunity to view the lesson in its entirety and identify the most salient segments.

These affordances of a whole lesson also come with some constraints. Having more video means that it takes more time to watch and find noteworthy aspects of a lesson; the intellectual demand increases after collection because a larger amount of data must be analyzed. Some teachers comment that, after they capture a lesson, they often have ideas about what in the lesson is worthwhile and can focus their viewing on these segments. That approach, however, minimizes the affordances of capturing a whole lesson, as a careful viewing of the entire lesson may reveal noteworthy interactions that the teacher had not been aware of during instruction. A whole lesson also requires more space to store and share the video, although this is becoming less of an issue as storage devices become smaller and have increased capacity. However, if groups of teachers are sharing video, consideration needs to be given about where to store whole lessons so they are easily accessible and minimize challenges associated with Internet speed and data transfer. Finally, although recording an entire lesson captures all of its richness, it can also detract from focusing on specific features of classroom interactions. There is so much to see in a lesson that it can be overwhelming to focus on particular features of interest, requiring more time and attention to review and identify relevant segments. In reference to reviewing a whole class video, one teacher commented, “I spent more time watching the video early on because you’re back and forth, back and forth, was that it? . . . I spent less time at the end because I knew what was going on.”

Selected excerpts. An alternative to taping a whole lesson is to capture selected excerpts. As is the case with capturing whole lessons, capturing excerpts provides both

Table 3
Affordances and Limitations of Capturing a Whole Lesson

<table>
<thead>
<tr>
<th>Issues</th>
<th>Affordance</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection</td>
<td>Minimal intellectual preparation</td>
<td>Requires more space on camera to capture</td>
</tr>
<tr>
<td>Analysis</td>
<td>Deciding what to attend to occurs after collection</td>
<td>Quantity of data to analyze can be overwhelming</td>
</tr>
<tr>
<td>Context</td>
<td>Context is more apparent, allowing for more nuanced analysis</td>
<td>Context can detract from the intended focus</td>
</tr>
<tr>
<td>Storing and sharing</td>
<td>Allows colleagues to see lessons from others’ classrooms</td>
<td>Requires more space to store and is challenging to share quickly with others</td>
</tr>
</tbody>
</table>
affordances and constraints (see Table 4). One affordance is that defining the types of excerpts one wants to capture ahead of time drives the focus of video capture. In particular, it allows teachers to focus on something other than what they usually focus on while teaching, which can give them different perspectives on their practice. We have worked with teachers to identify particular types of interactions to capture, allowing them to hone in on specific areas of interest. For example, if the goal for video analysis is to attend to student thinking, then teachers can capture videos of groups of students working together or whole class discussions that feature students sharing their work at the board. By defining an area of focus, fewer distractions arise from having too much data. Another affordance of capturing particular interactions is that the storage and sharing demands decrease—the video takes up less space and takes less time to view. Finally, defining the focus prior to taping reduces the amount of work required to identify a segment to share.

Defining ahead of time what segments to capture has constraints as well. One is that this approach requires more planning and intellectual demand prior to video capture. Careful consideration needs to be given to what will be captured at particular points in the lesson, and the lesson needs to be designed and implemented to elicit and capture the issue of interest. For example, if the issue of interest is student mathematical interactions in small groups, decisions need to be made about how many and which small groups to capture. Moreover, by narrowing the focus of capture, the broader context of the lesson is lost, requiring the teacher to provide relevant contextual information and supplemental materials to support later viewing and analysis. In addition, the sense of being inside one’s classroom can often get lost when teachers capture brief portions of a lesson. Though brief segments simplify the process of selecting clips to share, they also constrain what is available to be seen. This can be an affordance if the clips represent the issue of interest. However, if the issue of interest is not represented in the clips, then the limited scope of the segments restricts what is available for analysis.

Wearable cameras offer some unique affordances and constraints concerning the capture of selected excerpts of video. First, many wearable cameras offer selective archiving capability, which enables teachers to capture an event immediately after it occurs (see Sherin, Russ & Colestock, 2011, for further discussion). In addition, some allow teachers to mark moments while capturing an entire lesson, making it easier to return to selected segments. Choosing what to capture while teaching decreases the time needed to review an entire lesson at a later point. Moreover, selected clips are usually fairly brief, only a few minutes in length, requiring less time to review the clips and decide which segments to share. This also decreases the space needed to save, store, and share the video.

One constraint of wearable cameras, however, is the need to coordinate what a teacher wants to capture with the use of the technology—a teacher may find that he or she did not activate the camera at the right time to record a specific event. Another constraint for teachers using a wearable camera is that they need to monitor what they want to capture while they are also attending to the students, their learning, and the progression of the lesson, adding an additional challenge to the already cognitively complex work of teaching. We suspect that as teachers more frequently use this technology for self-capture and develop greater insight into the kinds of interactions that are worth capturing, these issues become of less concern. Finally, much of the classroom context can get lost capturing with this camera type; however, viewing several shorter segments from a lesson can allow viewers to seam them together to get a broad sense of the lesson and classroom as a whole.

### Deciding When to Specify the Tasks

Although there are many issues associated with the specification of tasks around self-captured video—including

<table>
<thead>
<tr>
<th>Issues</th>
<th>Affordance</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection</td>
<td>Focuses on specific areas of interest</td>
<td>Requires intellectual effort to decide what to collect</td>
</tr>
<tr>
<td>Analysis</td>
<td>Minimizes distractions</td>
<td>Narrowly focuses teacher noticing</td>
</tr>
<tr>
<td>Context</td>
<td>Fewer distractions</td>
<td>Risk of minimizing context</td>
</tr>
<tr>
<td>Storing &amp; sharing</td>
<td>Requires limited space, can share clips easily, and takes less time to view</td>
<td>Lose feeling of being in someone else’s classroom</td>
</tr>
</tbody>
</table>

Table 4

**Affordances and Limitations of Capturing Selected Excerpts**
the nature of the tasks themselves—we limit our focus to issues related to when tasks are specified for teachers. Tasks around self-captured video can be defined prior to capture, after capturing the video but prior to viewing and discussing the video in a group context, or when teachers come together to view self-captured video. Decisions about when tasks are assigned have implications for teachers’ analyses of the videos and for facilitating discussion of the videos.

**Teacher analysis and reflection.** Variations in when the task is specified have important implications for how teachers use the video to reflect on and analyze their practice (see Table 5). One affordance of having a specified task prior to video collection is that teachers can plan a lesson to meet the specified criteria. They could, for example, be asked to plan a lesson that includes small group work or discussion around a rich mathematical task. Setting the task ahead of time increases the likelihood that the teacher will capture what is intended; however, it is no guarantee. As one teacher remarked, “Something I thought was going to lead to really meaningful class discussion, or something I thought would make a great video, didn’t always happen.” Capturing something specific may become even more challenging if it requires the teacher to make in-the-moment decisions about the selection of video excerpts, as previously described.

The task for using self-captured video can also be specified after capturing the video but prior to discussing it. At this stage of the process, teachers might be asked to select video excerpts to share with their colleagues, to submit for an assignment, or to analyze using a particular framework. One affordance is that as teachers engage in such tasks, they have adequate time for individual analysis and can replay a video multiple times to locate clips or make sense of what they are seeing. Another benefit of this approach is that it allows teachers to formulate their own ideas and impressions of what took place before hearing others’ ideas.

A teacher explained: “I think [discussing the video with others before analyzing it alone] would be really difficult. It would be really hard to discuss your deep thoughts on it. You wouldn’t have had time to think about it, so it would be hard to bring . . . your ideas to the table.” However, one constraint is that teachers may struggle with the task if they do not have a good understanding of the ideas they are studying or if they were unable to capture particular types of interactions in their video.

Specifying the task when a group comes together to view and analyze self-captured video typically involves teachers responding to a prompt immediately following the viewing. In this case, the group discussion emerges from the ideas that participants raise. An affordance of this approach is that the discussion is responsive to what teachers notice in the video and to their current thinking. This approach also allows teachers to work together to jointly make sense of the video, each contributing ideas to further elaborate the group's thinking.

However, this approach also has constraints. Teachers may take up unproductive ideas or issues that are peripheral to the issues of interest. Additionally, without sufficient time to analyze the video individually, some teachers may not be able to formulate their ideas fully, limiting the ideas that might otherwise be available for discussion. As one teacher explained, “If we were all to just analyze at the same time as a group, you probably wouldn’t have gotten as much diversity, especially in the discussion.”

**Facilitation.** When the task is specified also has implications for facilitation (see Table 6). Setting the task prior to capturing increases the likelihood that the video captured will allow the facilitator to work toward his or her goals for teacher learning. For example, if the facilitator wishes to focus on making sense of students’ mathematical thinking, teachers might be asked to capture instances in which students share their mathematical ideas. This requires, however, that teachers have some level of

### Table 5

**Teacher Reflection and Analysis Related to When a Task Is Specified**

<table>
<thead>
<tr>
<th>When the task is specified</th>
<th>Affordances</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior to video collection</td>
<td>Lessons can be planned around ideas of interest</td>
<td>In-the-moment analysis/selection while teaching a lesson</td>
</tr>
<tr>
<td>Post-video collection/pre-discussion</td>
<td>Time available for thoughtful individual analysis</td>
<td>Individual teachers may struggle to engage with the task</td>
</tr>
<tr>
<td>During discussion</td>
<td>Discussion is more responsive to what teachers notice</td>
<td>Provides limited time for teachers to formulate ideas fully and for group thinking to emerge</td>
</tr>
</tbody>
</table>

*Mathematics Teacher Educator* • Vol. 4, No. 1, September 2015
knowledge about the ideas under consideration; for instance, in the previous example, the facilitator must do enough advance work with teachers for them to know what counts as an example of student mathematical thinking (see, for example, Leatham, Peterson, Stockero, & Van Zoest, 2015). Setting the task early also provides the facilitator an opportunity to assess teachers’ initial understanding of particular teaching practices and ideas about which they are learning.

Specifying the task after video capture but prior to the group discussion offers the facilitator the possibility of intentionally using teachers’ analyses to guide the discussion. For example, a facilitator can set up the task where teachers capture a segment, view and provide initial analysis, and submit their initial ideas to the facilitator before the group comes together. In this way, the facilitator can see how teachers have analyzed a video on their own and then develop questions to elicit key issues during the group discussion. Alternatively, a facilitator might select key teacher-identified instances to discuss as a group. The facilitator might, for example, choose both teacher-identified examples and non-examples of an idea of interest to help teachers learn to differentiate among instances and thus develop a better understanding of the concept they are studying as it arises in practice. A constraint of such an approach is that the facilitator loses the ability to observe teachers’ initial reactions to the video.

The task for self-capture and analysis may also be defined between capture and when a group shares and discusses the video, without requiring the teachers and facilitators to view the clips in advance of the group discussion. In this case, the group has an awareness of what they will focus on when they come together to discuss the clips (e.g., classroom discourse or student mathematical thinking) but there is less advance preparation. In addition, the group has a sense of what types of interactions and events to attend to when they get together to view the videos and can have more focused discussions, without demanding much of their time before coming together. Constraints include more work on the part of the facilitator to maintain a focus on the ideas of interest and greater difficulty ascertaining ideas that individual teachers have about the video.

Finally, a facilitator might specify the video analysis task when a group comes together to view the video for the first time. In this case, an affordance is that the facilitator can gain insight into teachers’ initial reactions to the video and develop a good sense of what is most salient to them. A constraint of this approach is that it requires more in-the-moment facilitation, particularly if unexpected ideas surface or unproductive ideas take hold.

### Discussion

Planning for and using self-captured video raises a variety of issues. We focused on three that we consider central to productive use of self-captured video for mathematics teacher education and professional development: where to position the camera, how much video to capture, and when to specify the task. We discussed these issues independently, but we recognize that in practice there is much balancing to do among them. For example, specifying the task has implications for camera placement and how much video to capture. In particular, asking teachers to capture short excerpts requires some degree of in-the-moment analysis to position and operate the camera, something that can be difficult for teachers to focus on while they are teaching. At the same time, asking teachers to capture a whole lesson requires less planning ahead of time but also has implications for identifying a clip to share and for facilitation. Having a clear sense of the purpose for using video, particularly with respect to supporting mathematics teacher learning, will help guide decisions on these three issues. Other issues, such as the nature of tasks for using self-captured video, strategies for sharing and discussing video, and orchestrating meaningful discussions with video that teachers capture and share with one another, are also important to consider but are beyond the scope of this article. In the next section, we offer some recommendations related to these issues based on our experiences.
Recommendations

From our work with teachers, we have identified several strategies for making the most of self-captured video. We offer the following recommendations.

Make the purpose explicit. We strongly recommend making the purpose of capturing and using video clear to teachers from the start. We have found that as teachers develop a deeper understanding of the specific purpose for capturing video, they also develop a more nuanced understanding of the benefits and limitations of different camera positions in reference to that purpose. One way we have made the purpose explicit is through the use of frameworks that specify what teachers will be looking for in the video (e.g., van Es, Cashen, & Auger, 2011). For example, in teacher education settings, we have provided detailed frameworks to provide criteria that teachers will identify in the classroom video when they occur, such as productive student thinking (e.g., Stockero, Peterson, Leatham, & Van Zoest, 2014) or evidence of maintaining high cognitive demand of student tasks (e.g., Stein & Smith, 2011). Providing teachers with such frameworks gives them a sense of the purpose for the video capture, allowing them to make informed decisions about which portions of the lesson to capture and how to position equipment to ensure they are capturing video that will allow them to apply the framework. One teacher we worked with explained,

[The goal] is something that needs to be made very clear from the get go and needs to be reminded every step of the way—that we aren’t just looking to have conversations about teaching and learning. We are specifically looking to discuss student thinking.

Provide multiple opportunities for video capture. A second recommendation is to allow teachers multiple opportunities to self-capture video. We recommend this for several reasons. First, positioning the camera and/or microphone can be difficult; teachers benefit from having some practice—trying out different placements, watching the video, and reflecting on the benefits and drawbacks of these placements. One teacher shared several issues and concerns that arose for her when she videotaped in her classroom, “What to tape and whether I would get anything good . . . where should I put [the camera], whether I should put it in the front of the class during a discussion or try to do a group . . . which group do I pick.” Second, when capturing short video clips, we have found that it takes time for some teachers to remember to activate a camera while they are also teaching a lesson. Self-capturing video while teaching seems to become easier with experience. We find it particularly helpful to require teachers to first capture video where the stakes are low (i.e., when it is not imperative that the video be of high quality for the intended purpose) in order to become proficient with video capture before they do so under the pressure of needing to capture high-quality video for a specific purpose.

Support teachers in designing and enacting lessons that include events of interest. Finally, what is captured on video is what is available to be seen, analyzed, and discussed. If the purpose of video capture and analysis is to examine student thinking, for example, then attention needs to be given to eliciting student thinking during instruction. Likewise, if the goal is to capture student-to-student discourse, then the lesson needs to provide opportunities for students to communicate with one another. This may require supporting teachers in designing and learning to enact lessons that will afford opportunities for particular types of events and interactions to arise during instruction. Our experience has been that through multiple cycles of video capture and analysis, teachers come to see that particular types of interactions may not be occurring in their classrooms, and they then attempt to create such opportunities in future lessons. This further supports our prior recommendation for multiple opportunities to capture video. Not only do teachers develop routines and practices for capturing and selecting clips, but they also come to develop a better sense of the kinds of clips that lead to productive discussion and allow them to gain deeper insight into the complexities of teaching and learning.

Conclusion

We conclude by proposing that the act of self-capturing video in and of itself is a form of professional development. By taking on this role, teachers become more deliberate and intentional in two important and mutually supportive ways: (1) planning for and enacting particular types of interactions in their classrooms and (2) planning to capture these interactions for later viewing and analysis. As a result, teachers become authors of their own learning. Teachers’ active participation in self-capture gives them more agency over what they capture and share with colleagues, how they analyze and reflect on their own and others’ teaching, and how they collaborate with others to explore the particulars that arise in their instruction. We conjecture that it is this ability to place the learning in teachers’ hands that makes self-captured video particularly promising for improving mathematics teacher education and professional development and, more important, mathematics teaching. Finally, we recognize that it can often seem daunting to capture video of one’s own teaching and share
this video with others and to navigate the logistical and substantive issues related to using self-captured video. We hope, however, that we have provided some guidance in getting started using this tool by specifying several areas in which teachers and teacher educators can make deliberate choices about how to make the most of teacher self-captured video.

References


Making the Most of Teacher Self-Captured Video


Authors

Elizabeth A. van Es, University of California, Irvine, 3455 Education, Irvine, CA 92697-5500; evanes@uci.edu

Shari L. Stockero, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931; stockero@mtu.edu

Miriam G. Sherin, Northwestern University, 2120 Campus Drive, Room 313, School of Education and Social Policy, Annenberg Hall, Evanston, IL 60208-0001; msherin@northwestern.edu

Laura R. Van Zoest, Western Michigan University, Department of Mathematics 5248, Kalamazoo, MI 49008-5248; laura.vanzoest@wmich.edu

Elizabeth Dyer, Northwestern University, 2120 Campus Drive, Evanston, IL 60208; ElizabethDyer2015@u.northwestern.edu

Mathematics Teacher Educator • Vol. 4, No. 1, September 2015
Appendix A: Equipment for Teacher Self-Captured Video

A range of equipment is available for teachers to use to self-capture video of their teaching. When choosing equipment, we find it useful to consider how easy the camera is to use, how easy it is to upload video from the camera, and the quality of the video and audio produced. Teachers may prefer to use cameras with which they are already familiar, whether that means a traditional video camera or their personal smart phone. In any case, becoming familiar with the range of equipment that is available can help teachers get the most out of their recording experience. We also provide suggestions on cameras and accessories to get started.

Video Cameras

We have worked with teachers using four different types of cameras to self-capture video: conventional video cameras, pocket cameras, smartphones/tablets, and wearable cameras (see Table A). Conventional video cameras have been used to record classrooms for many years. Current versions are much smaller than their predecessors and offer digital recording options. Though these cameras can be held and carried around the classroom, teachers typically place them on a tripod and leave them stationary while videotaping. Conventional video cameras usually offer the best picture quality for the price, and offer optical zoom, which permits close-ups without degrading picture quality. Many of these cameras record to removable memory cards so teachers can select the amount of storage that is appropriate, although some models only come with built-in storage capabilities.

Pocket cameras are similar to conventional video cameras in their features, although they have reduced optical zoom capabilities, if any at all. Compared to conventional cameras, the quality of the image may not be as clear when using the zoom capability. However, these cameras are small enough that teachers easily carry them around the classroom and record selected aspects of a lesson. Their small size also allows them to be placed on desks with groups of students with minimal disruption.

Phones and tablets are a relatively new way to capture video that can be particularly convenient if teachers already own one of these devices. One benefit of smartphones and tablets over the pocket camera is that the screen tends to be larger, which improves the capture and re-viewing experience. Many of these devices also offer simple editing tools to create clips from longer video or make other adjustments. However, these devices often do not accept removable memory cards, so video recording is limited to the free disk space on the device. Additionally, using an external microphone typically requires purchasing specific kinds of adapters or microphones.

Another camera option is the wearable camera (e.g., Drift Ghost-S, GoPro). These are commonly used to videotape sports and action from the point of view of a person. Although not a common choice for videotaping classrooms, some teachers prefer the perspective they offer. Some models (e.g., Drift Ghost-S, Looxcie HD) offer selective archiving, the ability to select moments of video to save immediately after they occur, or video tagging, the ability to mark moments while capturing video. This can be of interest to teachers who want to decide in the moment of instruction what video to capture. In addition, these cameras typically come equipped with a wide-angle lens; thus, the entire classroom is visible in the field of view from many positions. This perspective is often impossible to achieve in any position with the other types of cameras. Last, some models of wearable cameras also have the capability to be placed in a stationary position using a tripod, offering additional flexibility in their use.

Table A highlights the features we find noteworthy for each type of camera that are particularly common of high-quality cameras in that category. In addition, we identify specific cameras for each category as a starting point for identifying products to get started with self-capturing video.

Accessories

Two accessories frequently used to capture video are tripods and microphones. Tripods provide a more stable video image that can be easier to watch than video collected by someone who is moving. In the past, teachers have typically used a traditional tripod that stands on the floor and adjusts to different heights, allowing them to position the camera to capture the video from the height of the students. More recent options include desktop and magnetic tripods. The benefit of these newer tripods is that they are smaller and can be moved around a classroom easily. Magnetic tripods, in particular, easily attach to many surfaces in the classroom, such as whiteboards and metal file cabinets.

Additionally, external microphones are often used to increase the audio quality and provide flexibility to capture sound from a variety of classroom locations.
Microphones can be placed directly on the camera or they can also be placed in a different location than the camera by using an extension cord or a wireless microphone. To use an external microphone with a camera, the camera needs to have an external microphone jack, which may be a factor in a teacher’s camera selection. Many wearable cameras and conventional cameras now come with external microphone jacks, but they are rarely found on pocket cameras.

Memory cards and storage devices may also be needed for self-captured video. Most cameras use either a memory card to store the video or have built-in storage. Memory cards that are 16 GB or larger are recommended because they hold up to 2 hours of 1080p HD video. If teachers will be capturing extensive amounts of video over time, we recommend purchasing an external hard drive to store and organize the teacher-captured videos. Alternatively, various web sites...
(e.g., Dropbox®.com) exist for uploading, storing, and sharing video files.

**Video Storage and Sharing**

For most digital cameras, transferring video from a camera to storage has become fairly easy. In many cases, the camera can be connected directly to a computer and the files from the camera can be copied and pasted directly from the camera to a computer or external hard drive. When using removable memory cards, the card can either be inserted directly into the computer to download video and photos or an adapter can be plugged into the computer that allows users to upload images from memory cards. Smartphones and tablets, as well as some newer conventional and wearable cameras (e.g., GoPro Hero3), also have the option of sharing or uploading video directly from the camera over WiFi.

*(Return to page 7)*
Mentor-Guided Lesson Study as a Tool to Support Learning in Field Experiences*

Kristen N. Bieda, Jillian Cavanna, and Xueying Ji
Michigan State University

Field experience can be a rich site for intern teachers to develop the knowledge and skills they need for effective teaching. Lesson study has been shown to be a powerful form of professional development that enhances practicing teachers’ mathematical knowledge for teaching through collaborative inquiry with their peers. In this article, we discuss the use of mentor-guided lesson study to support mentor and intern collaboration in the field and share what we have learned about its potential to support interns’ attention to student thinking. We will also share insights from the field for those interested in implementing this activity in teacher preparation coursework.

Key words: Preservice teacher preparation–secondary; Lesson study; Mathematical knowledge for teaching

One common feature of both traditional and alternative teacher certification programs is the experiential learning structure known as “student teaching,” “field experience,” or “internship.” Regardless of what this structure is called, the purpose of the apprenticeship experience is for a novice teacher to take on increasing responsibility for instructing learners in actual classroom environments with mentoring from a more experienced teacher. The National Council on Teacher Quality reports more than 1,400 institutions of higher education require completion of a student teaching experience for teacher candidates (Greenberg, Pomerance, & Walsh, 2011). Moreover, teachers cite the field experience as the most valuable part of teacher preparation (Lortie, 1975).

The field experience provides prospective teachers with opportunities not only to engage in supervised practice but also to observe practice. How well prospective teachers learn by observing practice, however, is unclear (Brophy, 2004); in particular, novices may not yet have skills to notice key features of classroom instruction to support their learning (Star & Strickland, 2008). Given the recent attention to the role of mathematical knowledge for teaching (MKT) needed for mathematics teachers to engage in what Lampert, Beasley, Ghousseini, Kazemi, and Franke (2010) call ambitious mathematics teaching, we were interested in investigating whether structures supporting collaboration between prospective teachers and mentor teachers could focus prospective teachers’ attention on student thinking during classroom instruction and therefore support their development of MKT.

Recent research has examined the effects of action research on prospective and mentor teachers (Levin & Rock, 2003). Lesson study, a particular form of action research, has been shown to develop teachers’ MKT (Fernandez, 2005). Over the past 2 years, we have developed an assignment called mentor-guided lesson study for our secondary mathematics pedagogy courses to encourage collaboration between mentors and prospective secondary teachers (PSTs) and to focus prospective teachers on salient features of classroom instruction such as student thinking, thus heightening the knowledge and skills candidates develop in early field experiences. In this article, we will share findings from a mixed-methods study exploring the efficacy of the mentor-guided lesson study assignment in achieving these goals. Specifically, we will discuss: To what extent does mentor-guided lesson study support PSTs in noticing features of instruction relevant to developing knowledge for mathematics teaching?

Mentor-Guided Lesson Study

Initially used in Japan and widely used in other countries, lesson study is a form of teacher professional development that has gained increasing popularity in the United States (see Yoshida, 1999, and Stigler & Hiebert, 1999). In traditional lesson study, a team of teachers collaborates through four phases: (1) they formulate a goal for student learning; (2) they study their curriculum and plan a lesson that attempts to address the goal; (3) they investigate the implementation of the lesson by having one member of the team teach the lesson while the other members collect data through observation; and (4) they revise and

* In accordance with MTE policy regarding conflicts of interest with the editor, the review process for this manuscript was handled by Lynn Stallings, Kennesaw State University. This article was submitted and accepted under the editorship of Margaret Smith, prior to the first author’s term as associate editor of the journal.
reteach the lesson based upon careful reflection on the observations of the initial lesson implementation. Mentor-guided lesson study is a variation of lesson study in which the team includes a pair of PSTs and their mentor teacher. At our institution, PSTs complete a weekly 4-hour field experience as part of a secondary mathematics methods course required prior to student teaching or internship. The frequency with which PSTs visit their placements varies based on their schedules. Some students may visit schools just once a week, while others split their hours over multiple days. Mentor-guided lesson study (MGLS) is an assignment integrated with the field experience; it is PSTs’ primary opportunity prior to student teaching to teach full lessons in the field.

Each cycle of the MGLS activity consists of four phases (goal setting, planning, teaching, and reflecting), and teams are required to complete two cycles during the semester (see Table 1). The role of teacher of the study research lesson is rotated among the team members to support collaboration. The mentor teacher assumes the role of teacher in Cycle 1, and the PSTs co-teach the research lesson in Cycle 2. Within a cycle, the PSTs complete five online Collaborative Learning Logs, which consist of reflection prompts designed to focus the team’s discussions in particular ways during each phase. Sample questions from the Collaborative Learning Logs are provided as we discuss each phase of the MGLS activity below.

Table 1
Breakdown of Assignments During Mentor-Guided Lesson Study Phases

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle 1</td>
<td>Goals Development Log</td>
<td>Topic Study Log</td>
<td>Lead by MT</td>
<td>Post-Lesson Reflection Log</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observation Guide Log</td>
<td></td>
<td>Post-Lesson Debrief Log</td>
</tr>
<tr>
<td>Cycle 2</td>
<td>Goals Development Log</td>
<td>Topic Study Log</td>
<td>Lead by PST-Ms</td>
<td>Post-Lesson Reflection Log</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observation Guide Log</td>
<td></td>
<td>Post-Lesson Debrief Log</td>
</tr>
</tbody>
</table>

* Indicates places during the assignment phases where the course instructor provided feedback.

The prompts for the online logs were developed from materials in Lesson Study in Practice: A Mathematics Staff Development Course (Gorman, Mark, & Nikula, 2010).
In the Planning phase, the team designs a lesson plan to address their goal(s) and decides how team members will observe and document students’ thinking during the lesson. As in the Setting Goals phase, a five-prompt online log (Topic Study Log) guides the PSTs in thinking about the mathematical content of their research lesson prior to meetings with their mentor teacher to plan the research lesson. Another four-prompt log (Observation Guide Log) structures teams’ planning for data collection and observation. Prompts in these planning logs include—

“Discuss how you have addressed your concerns about your understanding of the topic. This may include working through related problems with your mentor teacher or placement partner or reading through textbook materials.” (Topic Study)

“In what ways did the lesson seem effective (or ineffective) in helping students understand the main mathematical ideas in the lesson?”

After the postlesson discussion, the PSTs complete a final, two-prompt online Post-Lesson Reflection Log. The intent of this log is not only for PSTs to report on any modifications their team decided to make to the research lesson, but also for participants to reflect upon their team’s collaboration. Prompts on the Post-Lesson Reflection Log include—

“Discuss how your team came to make decisions about the revisions. You will want to discuss what seemed to be the most important pieces of evidence collected about what students learned.”

“Discuss your impressions of how well your team collaborated during the debrief and revision process.”

PSTs also submit a revised lesson plan and reflection paper at the end of the cycle. The process of completing this final phase is consistent across both cycles. Cycle 1 and Cycle 2 occurred approximately one month apart, based on the curriculum schedule of the mentor teacher’s class. The teams were not required to reteach revised lessons within cycles because of typical scheduling issues involved with attending the placement on nonconsecutive days.

In addition to these logs, each PST pair submits a completed lesson plan to the course instructor in advance of the lesson. For Cycle 1, the PSTs construct the lesson plan in a format used throughout our program. Because the mentor teacher assumes the role of teacher of the research lesson in Cycle 1, the course instructor does not provide feedback on the lesson plans to honor the mentor teacher’s expertise and sense of efficacy in planning lessons. In Cycle 2, when the PSTs assumed the role of teacher, the course instructor provided written feedback about potential issues in lesson implementation, suggests student responses that may have been unanticipated, and supports PSTs’ skills in lesson planning.

In phase three, Teaching, the PSTs observe their mentor teachers teach the research lesson in Cycle 1, whereas the PSTs take the lead and co-teach the lesson while the mentor teacher and, in some cases, other interns, observe in Cycle 2. Although not required, some of our teams were able to video record the research lesson enactment.

In the fourth and final phase of each cycle, the Post-Lesson Discussion, teams reflect upon the research lesson enactment and on possible modifications that would better achieve the desired learning goals. Within 24 hours of the lesson enactment, each PST completes an online, four-prompt Lesson Reflection Log. This log consists of prompts such as—

“List one or two observations you would like to share with the team. Be as specific as possible about the evidence of student thinking that you observed.”

“Discuss your team’s decision on possible modifications that would better achieve the desired learning goals.”

To assess PSTs’ work in the lesson study cycles, the course instructor checked all of the log responses for completion and used a rubric to evaluate the initial lesson plan, the revised lesson plan, and the final reflection paper.

**Mentor-Guided Lesson Study in the Context of Secondary Mathematics Methods Courses**

We implemented MGLS in the second course (which we will call Methods II) of a four-course sequence of mathematics methods courses taken by prospective secondary mathematics teachers at a large university in the Midwestern United States, where students earn a bachelor’s degree in mathematics while taking education credits required for certification. Teacher candidates typically take Methods I in fall semester, followed by Methods II in the spring prior to a year-long, full-time student teaching internship. Methods III and IV are taken during the student teaching internship year. At the time of this investigation, 17 students were enrolled in Methods II, and the first author was the instructor for both Methods I and II. During the Methods I course, 17 PSTs were grouped with eight mentors in four different schools, for a total of eight MGLS teams. Pairs (and one triad) were required to be in their placements for at least 4 hours per week throughout Methods I and II.
In addition to the field placement, Methods II included a seminar for 3 hours per week, where students explored practices supporting inquiry-oriented mathematics instruction, and a lab where students learned about practices related to teaching students with special learning needs. The MGLS was a focal assignment in the Methods II course and was designed to deepen PSTs’ learning and capacity to collaborate with mentor teachers. The first MGLS cycle occurred after PSTs had been working in their mentor teacher’s classroom for approximately two to three months in the fall semester, so that the PSTs were familiar with the students, curriculum, school context, and their mentor teachers.

Although all of the mentor teachers had taught for more than 5 years, mentors’ experience with lesson study varied. Three of the mentor teachers were first-time mentors for PSTs in this program and had not previously participated in lesson study. Three other mentors had participated in MGLS once in the past, so they were familiar with the assignment. The two remaining mentors had participated in MGLS once and had completed three lesson study cycles with colleagues. Although these two mentors had lesson study experience, they had not used the Collaborative Learning Logs to guide each phase of the cycle.

Regardless of experience level with lesson study, all mentors participated in a half-day workshop designed and facilitated by the authors before PSTs worked in field placements at the beginning of Methods II. During this half-day workshop, mentors learned about phases of lesson study and the various assessments, such as the lesson plan requirements and Collaborative Learning Logs that the prospective teachers were to complete during MGLS. Mentors also watched videos of a secondary mathematics lesson study team during a postlesson discussion (from Gorman, Mark, & Nikula, 2010) and completed a simulation in which they wrote and revised lesson study goals for a given standard with guidance from the Methods II course instructor.

Investigating the Nature of PSTs’ Learning through Mentor-Guided Lesson Study

Research on teacher learning during lesson study highlights how the process supports teachers’ learning about mathematics for teaching (Fernandez, 2005) and promotes rich conversations about mathematics (Parks, 2008). Because these opportunities were some of PSTs’ first to engage with students in an instructor role, we wondered which aspects of instruction PSTs would attend to as they completed each cycle of MGLS. To study the aspects of teaching and learning secondary mathematics foregrounded by PSTs during MGLS, we used techniques of grounded theory (Strauss & Corbin, 1990) to examine patterns and generate descriptions of themes present in the PSTs’ reflections. The analytic process that we used to generate codes and themes in our data is specified in the following section.

Analytic Process

The primary source of data for this investigation was the Collaborative Learning Logs, which offered identical prompts for all PSTs, unlike the teams’ lesson study conversations, which varied in content and form. Our analysis began with open coding, where we identified instances in the logs in which each of the PSTs reflected upon particular features of the research lesson related to his or her knowledge of mathematics, of students, and of teaching. After the initial open coding, we generated a complete set of subcodes for three emergent themes, which we named Analyzing Mathematics, Attending to Student Thinking, and Analyzing Teaching Moves. Using the constant comparative method (Strauss & Corbin, 1990), we further refined these categories and developed the coding scheme shown in Table 2.

The category Attending to Student Thinking differed from the other two categories, both of which have “analyzing” in their name, because these codes reflected observations, rather than evaluations or assessments, of student thinking. Reflections coded into Analyzing Teaching Moves were more evaluative in nature, such as discussing whether a teacher move was appropriate in relation to students’ engagement or the class discourse about the tasks assigned. Likewise, reflections coded into Analyzing Mathematics evaluated aspects of the students’ or teachers’ mathematical work in terms of its validity or worth in supporting conceptual thinking.

It is worth noting that particular prompts in the logs may have encouraged reflections related to one of the categories more than others. For instance, this prompt from the Lesson Reflection Log—“List one or two observations you would like to share with the team. Be as specific as possible about the evidence of student thinking that you observed.”—would likely result in responses coded as Attending to Students’ Thinking. However, further analysis of subcodes (listed in Table 2) assigned to those responses yielded more information about the quality of such reflections and whether PSTs’ stepping into the role of lead teacher in Cycle 2 had possible effects upon what PSTs noticed about teaching and learning during the lesson study. In the sections that follow, we describe the categories in more detail and provide examples from the data to illustrate particular subcodes.

Analyzing mathematics. Statements in this category include PSTs’ focus on the mathematics of a particular
research lesson, such as the mathematical features of students’ statements or written responses or the mathematics of teaching moves. None of the statements coded to this category included pedagogical features of the lesson. The following are examples from the category Analyzing Mathematics Related to Students’ Thinking:

“No students discussed all but one possible method to estimate the area.”

“The teacher graphed both the function and inverse function on the same set of axes, but this is mathematically incorrect.”

The first example specifically analyzes the mathematics content of student responses, not the students’ interaction or instruction. In the second response, the focus is on the mathematical content of the instructional move, not its pedagogical aspects.

**Attending to Student Thinking.** This category includes observations PSTs made about students’ thinking, procedures, or interactions, such as how students made sense of mathematical concepts presented during the research lesson as well as any errors. The subcategories of this code distinguish between a focus on students’ understanding of concepts and students’ ability to execute procedures. Examples of responses coded to some of these categories include—

“No students used a chart or graph to help them understand the properties of the equation.” (Observations of Student Understanding)

“Students were having trouble determining the effect of the negative leading coefficient on the shape of the graph.” (Observations of Student Difficulties or Errors)

We distinguish between Observations of Student Understanding and Student Difficulties or Errors by analyzing the semantics of the statements: If the statement about students’ thinking was framed in terms of what students know or understand, it was coded as Observations of Student Understanding. In the second example, the comment was about something that students were struggling to understand. We coded these as Observations of Student Difficulties or Errors.

**Analyzing teaching moves.** This category captures PSTs’ statements related to their own teaching moves or those of their partner PST or mentor teacher. This category included analysis of the reasoning for a particular teaching move. The subcategories group the PSTs’ observations based on what they interpreted as the motivation of the teaching move. Some examples include—

“Since students were having problems plotting points, then the lesson before should have covered how to construct a graph.” (Teaching Moves in Response to Student Difficulties)

“The exit slips helped us see exactly how the students were thinking and what they thought was the best approximation.” (Teaching Moves Related to Student Thinking)

“My accommodations of group work and reading out loud had a positive effect with the student I included it for. For instance, he spoke up during class discussion and was talking openly during group work.” (Teaching Moves Related to Student Engagement)
Data Collection and Analysis

The results in this paper are based on our analysis of PSTs’ responses to the Lesson Reflection and Post-Lesson Discussion Collaborative Learning Logs. We focused on these logs because they represent PSTs’ reflections on the research lessons. We examined similarities and differences in the coding (see Table 1) assigned to PST responses on logs across Cycle 1 and Cycle 2 to ascertain the potential influence of the shift in roles between the PSTs and the mentor teachers. We translated the Google Form spreadsheet entries from the two logs into de-identified text (.txt) files and coded them using the software HyperResearch (version 2.8.3, www.researchware.com).

Results

In the following section, we present results of our analyses indicating, importantly, that MGLS supports PSTs in attending to, and reflecting upon, students’ thinking. We also discuss how the mentor’s lesson study experience may directly influence the focus of PSTs’ reflections in the Lesson Reflection and Post-Lesson Discussion Logs. We present these findings in two ways: (1) general quantitative results from our codes of the logs and (2) detailed descriptions of specific trends noted in the quantitative findings.

Part I: Overall Results

We coded 340 instances of the three categories of reflection (see Table 2) from the 17 participants’ Lesson Reflection and Post-Lesson Discussion Logs from Cycle 1 and Cycle 2. Figure 1 summarizes the frequency of codes across the three general categories. One clear finding is that instances of Attending to Student Thinking occurred more frequently than other two categories, and that Cycle 2 included more codes overall than Cycle 1. Although the mean number of instances of Attending to Student Thinking was greater in Cycle 2 than in Cycle 1 (141 versus 91, respectively), a paired sample t-test showed this difference to be insignificant (p = 0.162).

The frequency of Attending to Student Thinking codes far outnumbered the frequency of codes for Analyzing Teaching Moves, which is interesting because the responses coded came from logs written after the lesson enactment. This finding suggests that MGLS focuses PSTs’ observations on their students’ thinking. The relatively low frequency for the code Analyzing Mathematics was due, in part, to the fact that we are reporting findings from analyses of selected log responses (see Data Collection and Analysis section) and is likely not indicative of the PSTs’ lack of attention to the mathematics of the research lesson. Results from coding other logs, such as the Topic Study Log, would likely yield a different balance of categories emerging from the coding.

As for the increase in responses coded into one of the three general categories from Cycle 1 to Cycle 2, we propose several explanations. One such explanation is that the PSTs may have become more familiar with the prompts and the process of lesson study. Another possible explanation is the shift in the PSTs’ role from Cycle 1 to Cycle 2. Becoming the lead teachers in Cycle 2 may have afforded PSTs opportunities to recognize important moments in the lesson related to student thinking less apparent to an observer. The design of our study, particularly that all participants completed the cycles using the

Figure 1. Frequency of codes by cycle.
same protocol and shifted into the role of research lesson teacher in Cycle 2, does not allow for further investigation into these possibilities.

An initial focus of the work was to examine the mentor’s influence on the nature of PSTs’ reflections during the MGLS process. One initial hypothesis of the study was that the more prior experience a mentor had with the lesson study process, especially lesson study experience with other colleagues, the more likely his or her collaborations with PSTs would influence the sophistication of PSTs’ noticing during the lesson study process. Figure 2 shows mentors’ level of experience and frequency of codes across cycles, where mentors were assigned to one of three different experience levels. Mentors with no prior experience doing MGLS or any lesson study are categorized as “No Experience.” Mentors who had some prior experience with MGLS, but no experience in lesson study with inservice teacher colleagues, were categorized as “Some Experience.” Mentors who had prior experience both in MGLS and in lesson study with inservice teacher colleagues, were categorized as “Most Experience.” Results from a repeated measures ANOVA found level of experience to be a significant predictor of increased frequency in responses coded to one of the three categories of observation ($F(2,13) = 4.515, p = 0.032$).

In the following section, we examine these general findings on the nature of PSTs’ reflections on student thinking and the influence of mentor teachers’ prior experience with lesson study on those reflections so that we can better understand how PSTs’ experience in doing MGLS supported reflections to develop MKT.

**Part II: Zooming In on Reflections About Students’ Thinking**

One of the most frequently mentioned outcomes of lesson study is its ability to focus teachers’ attention on student thinking (e.g., Lewis, Perry, & Murata, 2006), and our findings showing this to be the case with MGLS are especially important in teacher education because they suggest that PSTs may learn more in field placements when engaging in an activity such as MGLS. The formal nature of observation in lesson study may account for some of this, but the tradition of discussing observations of student thinking before discussing revisions to the research lesson may also serve to focus teachers’ attention to student thinking. A closer look at the responses coded as Attending to Student Thinking revealed more about the salient aspects of students’ thinking for PSTs during the MGLS process, particularly as they transitioned from role of team member in Cycle 1 to the role of research lesson teacher in Cycle 2.

As Figure 3 shows, with the exception of mathematical procedures, all of the subcategories of Attending to Student Thinking increased from Cycle 1 to Cycle 2. Perhaps more important, participants noted aspects of engagement and understanding more than the other categories. To better understand why participants noticed these aspects more than others in this category, we investigated a potential relationship between mentors’ level of experience with lesson study and the frequency of particular categories assigned to responses related to Attending to Student Thinking. Figure 4 shows these results.
One noticeable finding is that responses coded as Attending to Student Engagement were predominantly from PSTs whose mentors had no prior MGLS experience. It is unlikely that this finding reflects particular differences in these PSTs’ knowledge and skills compared to other members of their cohort, as pairs of PSTs were randomly placed with a mentor depending upon factors such as schedule availability. Further, we cannot assume that more reflection about student engagement means that student engagement is a problematic aspect of a mentor teacher’s classroom. It is more likely that the differences in the teaching practice of mentors who have lesson study experience and those who do not contributed to the differences in PSTs’ noticing, because prior lesson study experience likely contributes to the mentors’ habits of reflection and noticing. Another relevant factor may be that the PSTs’ role in discussions between themselves and their mentor teachers could account for PSTs’ reflection upon certain aspects of the research lesson.

A deeper look into patterns among the reflections coded as Attending to Student Thinking revealed distinctions related to the specificity of the reflections. van Es and Sherin (2008) argue that specificity is one characteristic of increasingly more sophisticated teacher noticing. A high degree of specificity would help all team members understand what they observed, and contributed to a team’s ability to modify the research lesson to address students’ thinking.

The following examples illustrate differences between high and low degrees of specificity. These responses were all coded as Attending to Student Engagement:

“Without question, the most telling observation for this entire lesson came during 2nd hour, about 10 minutes into the group work time. At this time, I noted that 22/24 students in the room were either standing at the board writing/discussing with their group, or were at their desk actively calculating and crunching numbers.” (Cycle 2, Lesson Reflection Log, mentor had most experience with lesson study)

“This was interesting because the teacher had gone over an example of what to look for and they seemed to not pay attention or see the correlation between the example they had done and the explore task.” (Cycle 1, Lesson Reflection Log, mentor had no experience with lesson study)

In the first example, the PST-M provides exact counts of students doing particular actions, whereas the second example simply states “there were a lot more people off task.” Highly specific reflections about student engagement provided precise counts for how many students were engaged or disengaged and specified actions that students were doing or saying as evidence of engagement. Given the context of MGLS, a high degree of specificity indicates either that individual PSTs carefully accounted for

![Figure 3. Frequency of subcodes for Attending to Student Thinking across cycles.](image-url)
episodes during the research lesson or that the conversations between the PSTs, or as a team with their mentor teacher, referenced detailed actions made by the students and/or teacher.

The majority of responses coded to the Attending to Student Engagement category had a low degree of specificity (see Figure 5). This finding is unsurprising given that the construct of Attending to Student Engagement involves noticing student behavior at a relatively large grain size. In contrast, the construct of Attending to Student Understanding involves noticing student cognition and defining what mathematical skills or ideas students understand. When analyzing level of specificity for responses coded as Attending to Student Understanding, our findings contrasted those from our analyses of Attending to Student Engagement. We found responses coded to Attending to Student Understanding were more likely to be highly specific, especially for teams whose mentors had at least some experience with MGLS (see Figure 6).

The following two examples illustrate the differences between responses with high and low specificity for reflections coded as Attention to Student Understanding:

“Students made some really good observations on the question ‘How does a student make it to Box A? B? . . . G? How many heads or tails would they have to get? Does order matter?’ Two students only listed one possibility for each and said that there are more possibilities in the middle. Another student noticed that order did not matter as long as you had a certain number of heads or tails you would end at a certain spot.” (Cycle 2, Lesson Reflection Log, mentor had most experience with lesson study)

“What I mean is a lot of students individually were getting stuck on the questions and from each other’s help in the group and some probing from the teacher, I think they got a lot more out of the word problems. The students strengthen their skills in pulling and gathering information.” (Cycle 2, Lesson Reflection Log, mentor had no experience with lesson study)

Both responses describe students’ understanding during the lesson, but they vary in terms of how specifically they identify students’ mathematical thinking. The response from a PST working with a mentor who had the most experience with lesson study articulates specific student responses, whereas the PST-M working with a mentor who had no prior experience with lesson study only provided
“I think they got a lot more out of the word problems” instead of stating specific responses from students or mathematical ideas students appeared to be learning while solving the word problems. We acknowledge that mentors’ level of lesson study experience only correlates with increases in specificity of PSTs’ reflections during MGLS, but our findings raise the possibility that implementing MGLS is particularly successful when mentors have some prior experience with lesson study.
What We Have Learned

These results suggest that the MGLS experience focuses PSTs’ attention upon student thinking during research lessons and addresses what Star and Strickland (2008), among others, acknowledge as a need to develop PSTs’ skills in noticing key features of classroom instruction. In particular, what PSTs notice during MGLS may support the development of knowledge of content and students (Ball, Thames, & Phelps, 2008). The results also suggest that PSTs’ capacity for noticing aspects of instruction related to students’ thinking and key teaching moves increases in Cycle 2 with more experience in doing MGLS. Levin, Hammer, and Coffey (2009), arguing against stage-based notions of prospective teacher development, show that it is possible, and desirable, for prospective teachers to attend to student thinking. From our work, we have found that MGLS is a promising means for promoting prospective teachers’ skill in attending to student thinking in ways that go beyond attending to students’ engagement.

As we have discussed, a likely factor influencing what PSTs reflect upon during MGLS is the mentor’s teaching practice. One explanation for why PSTs placed with mentors who have at least some experience with lesson study have a higher frequency of highly specific reflections coded to categories such as Attending to Student Understanding is that mentors’ prior experience with lesson study has shaped their practice in ways that prioritize building upon students’ thinking. In addition, it is plausible that the nature of the tasks in the research lessons and their cognitive demand is influenced by mentors’ prior experience with lesson study (and all forms of PD). The nature of the tasks in the research lessons would certainly influence what teachers have access to notice about student thinking. Moreover, the mentor’s ability to shift his or her practice in these ways is also likely correlated with MKT.

We caution against an interpretation of our findings that PSTs guided by mentors inexperienced with lesson study are not noticing important aspects of teaching mathematics or developing MKT through MGLS. It may also be the case that mentors do not have as much influence in the quality of PSTs’ reflections. It may be the two-cycle structure of the assignment where PSTs take on the role of lead teaching in the second cycle that has the most influence on the quality of PSTs’ reflections. Our study design can only provide limited information about which aspects of MGLS were more influential in shaping PSTs’ reflections, as we did not systematically vary the conditions of the assignment.

Ideas for Supporting Successful Implementation of Mentor-Guided Lesson Study

The structure of lesson study, in any form, is not what promotes learning. Instead, it is the process of collaboratively discussing goals for student learning, mathematical tasks, evidence of student thinking, and teacher moves that enhances their professional knowledge. Our findings show that the MGLS helps PSTs attend to critical aspects of instruction, and our implementation of MGLS suggests two strategies that promote successful implementation of MGLS. First, we have found that the adoption of MGLS is most successful in classrooms where mentors agree to work with multiple PSTs and in schools where multiple mentors are paired with PSTs for MGLS. This arrangement encourages a community of practice around lesson study. Given that the activity occurs as an early field experience, where PSTs are placed in mentors’ classrooms for only 4 hours per week, it is feasible for experienced mentors to work with multiple PSTs.

Second, sharing materials such as the Collaborative Learning Logs is essential for supporting mentor teachers in facilitating discussions with PSTs about the lesson study throughout the cycle. Perry, Lewis, Friedkin, and Baker (2009) found evidence that toolkits containing suggested tasks and activities for teams of inservice teachers to complete during the lesson study process increased opportunities to learn MKT. The Collaborative Learning Logs we used were adapted from Lesson Study in Practice: A Mathematics Staff Development Course (Gorman, Mark, & Nikula, 2010), an excellent source for such materials. These resources support PSTs as they complete phases in the process, but they can also support the mentor teachers for when (or if) they decide to engage in lesson study with their colleagues.

Concluding Remarks

MGLS is grounded in the typical work of teaching, but it encourages participants to attend more deliberately to how the features of a lesson affect students’ thinking. Although MGLS is time intensive, the payoffs are significant. In engaging in the mentor-guided lesson study process, the PSTs are not just observing or assisting, but are also supported to reflect upon the key aspects of instruction highlighted in exemplary support for learning to teach (Feiman-Nemser, 2001). Anecdotally, mentors participating in the experience over the past 2 years report that the MGLS structure nudges them to be explicit about aspects of their practice they were not even aware were necessary to share with PSTs. It may be that important pieces of the canon of excellent teaching practice are so tacit that they cannot be taught through a textbook or observed in a video. Mentor-
guided lesson study experiences have the potential to support mentor teachers in accessing, and sharing, these aspects of practice to enrich what PSTs learn in early field experiences.

References


Acknowledgment

This work was funded, in part, by a Michigan State University Lilly Teaching Fellowship awarded to the first author.

Authors

Kristen N. Bieda, Michigan State University, 620 Farm Lane Road, 317 Erickson Hall, East Lansing, MI 48824; kbieda@msu.edu

Jillian Cavanna, Michigan State University, 354 Farm Lane Road, East Lansing, MI 48824; cavannaj@msu.edu

Xueying Ji, Michigan State University, 620 Farm Lane Road, East Lansing, MI 48824; jixueyin@msu.edu
Transforming Perceptions of Proof: A Four-Part Instructional Sequence*

Justin D. Boyle  
*University of Alabama

Sarah K. Bleiler  
Middle Tennessee State University

Sean P. Yee  
University of South Carolina

Yi-Yin (Winnie) Ko  
Indiana State University

Mathematics teachers are expected to engage their students in critiquing and constructing viable arguments. These classroom expectations are necessary, given that proof is a central mathematical activity. However, mathematics teachers have been provided limited opportunities as learners to construct arguments and critique the reasoning of others, and hence have developed perceptions of proof as an object that must follow a strict format. In this article, we describe a four-part instructional sequence designed to broaden and deepen teachers’ perception of the nature of proof. We analyzed participants’ reflections on the instructional sequence in order to gain insight into (a) the differences between this instructional sequence and participants’ previous proof learning opportunities and (b) the ways this activity was influential in transforming participants’ perceptions of proof. Participants’ previous learning experiences were focused on memorizing and reproducing textbook or instructor proofs, and our sequence was different because it actively and collaboratively engaged participants in constructing their own arguments, critiquing others’ reasoning, and creating criteria for what counts as proof. Participants found these activities transformative as they became more clear about what counts as proof, began to view proof as socially negotiated, and expanded their conception of proof beyond a rigid structure or format.

Key words: Reasoning and proof; Learning to prove; Constructing arguments; Critiquing arguments; Criteria for proof

Secondary mathematics teachers are expected to provide their students with opportunities to construct viable arguments and critique each other’s reasoning across all mathematics courses and topics (CCSSM, 2010; NCTM, 2000, 2009). These are noteworthy recommendations given the central role of proof in mathematics, and how engagement in reasoning and proving has the potential to deepen students’ understanding of mathematical concepts (Ball, Hoyles, Jahnke, & Movshovitz-Hadar, 2002; Hanna, 1995; Hersh, 1993). Furthermore, the treatment of proof in classrooms should be a negotiated communal activity that promotes understanding (Ball et al., 2002; Hanna, 1995; Hersh, 1993). To reach this end, G. Stylianides (2008) explains how classrooms can access deductive reasoning through a set of activities such as generating examples, identifying patterns, and making a generalization before constructing an argument. He explains these activities as the reasoning-and-proving framework that aligns with stages mathematicians pass through to produce a proof. While progress is underway (e.g. G. Stylianides & A. Stylianides, 2009; Karunakaran, Freeburn, Konuk & Arbaugh, 2014), more work is needed before a communally negotiated view of proof is materialized across secondary classrooms (Bieda, 2010; Furinghetti & Morselli, 2011; Steele & Rogers, 2012).

A current challenge with integrating proof as a central curricular activity is that based on prior experiences, undergraduate mathematics and mathematics education majors and practicing secondary mathematics teachers have developed narrow views and abilities with proof (Bleiler, Thompson, & Kračevski, 2014; Furinghetti & Morselli, 2011; Knuth, 2002a, 2002b; Kotelawala, 2009; Tabach et al., 2011). Mathematics teachers have been given limited opportunities as learners to construct arguments and critique the reasoning of others, and hence have developed narrow perceptions of proof. For example, some undergraduate students and practicing secondary mathematics teachers believe that proof must follow a strict format.

* In accordance with MTE policy regarding conflicts of interest with the editor, the review process for this manuscript was handled by Melissa D. Boston, Duquesne University. This article was submitted and accepted under the editorship of Margaret Smith.
Some practicing secondary mathematics teachers view proof as a topic of study in geometry (Knuth, 2002a; Kotelawala, 2009) or think that it is only appropriate for the most advanced high school students (Knuth, 2002a; Furinghetti & Morselli, 2011). Thus, teachers either present proofs as a final product or avoid proofs altogether (Furinghetti & Morselli, 2011), often believing that their students are unable to prove (Knuth, 2002a).

Instruction that focuses on presenting proof as a completed product limits learners’ access with how to construct and/or assess arguments, as students in these learning environments are passive participants (Hanna, 1995; Harel & Sowder, 1998; Solomon, 2006). Traditionally, the instructor or textbook serves as the sole producer and arbiter of proof in a classroom, which Harel and Rabin (2010) identify as the authoritative approach. Learning to construct proofs in an authoritative learning environment contributes to narrow views and limited understanding (Hanna, 1995, 2000; Harel & Sowder, 2007). We agreed with Harel and colleagues (Harel & Sowder, 1998; Harel & Rabin, 2010) that when students (K–16) experience proof through the authoritative perspective, they develop limiting views of proof and these views stay with them even after they become teachers. As we summarized below in Figure 1 (authoritative perspective), the beliefs that teachers harbor, as shown in the first trapezoid, toward the nature of proof in schools translates into classroom practices and assessment routines that lead to unproductive results. The student learning outcomes (i.e., students view proof as a rigid, formal object; students learn that they are unable to prove) are direct outputs of the classroom engagement and assessment routines where students view their role as passive consumers of presented proofs (Solomon, 2006). These student outcomes lead some educators to advocate for proof to become more reclusive, and others have called for it to be eliminated from curricula altogether (Hanna, 1995). To address this ongoing cycle grounded within the authoritative perspective, we find it essential to engage teachers in productive proof learning opportunities. The goal is to transform teachers’ perspectives away from the authoritative approach, so that proof has the potential to become a central part of the curricula.

We contend that if prospective and practicing teachers have opportunities to learn proof as a communal, negotiated, and sense-making process as recommended (Ball et al., 2002; Hanna, 2000), they will be better equipped to foster their students’ development of proof. In this article, we offer a general four-part instructional design that has the potential to transform learners’ perceptions of proof. We explain in detail our specific enactment of an instructional sequence and then provide a discussion of the key generalized features of the sequence that can be implemented across instructional contexts with different student populations. We seek to answer the following two research questions:

- In what ways do participants perceive this instructional sequence as different in relation to their prior experiences with mathematical proof?
- How (if at all) do participants perceive that the activities in this instructional sequence changed their perceptions of proof?

**Background**

The pervasive authoritative perspective directly conflicts with research recommendations regarding how to support students’ access and ability to construct proofs (Harel & Sowder, 2007; Hanna, 1995; Lannin, 2005; A. Stylianides, 2007a). The research indicates that students at all levels (K–16) should be provided opportunities to engage in proof as a process (e.g., generating examples, looking for patterns, and making a generalization) before developing a valid argument, as mathematicians do (Stylianides, 2008).
If students have a general idea about what is needed for an argument to count as proof, they can understand what they are working toward (Stylianides, 2007a, 2007b). The disconnect between research and classroom practice requires attention, and a productive way to address the disconnect is through changing students’ and teachers’ learning opportunities. We designed Figure 2 below to summarize our conceptualization of the research recommendations for proof, positioning proof as a communal activity. By contrasting Figure 1 and Figure 2, we can see how research recommendations (Figure 2) conflict with the authoritative perspective (Figure 1).

Starting with beliefs (as shown in Figure 2), students should be afforded opportunities to engage in proof as a process, even if their initial attempts to construct arguments are not proofs (G. Stylianides & A. Stylianides, 2009; NCTM, 2009, 2014). Teachers often believe they should focus on what students produce as correct (proof) or incorrect (nonproof) (Furinghetti & Morselli, 2011; Knuth, 2002a), as opposed to working from what students produce as part of a learning process. Lannin (2005) conducted a teaching experiment to develop the ability of 25 sixth-grade students to construct proofs. Lannin found that students mostly depended upon empirical arguments when they constructed arguments on their own. However, during the whole-class discussion, students were able to verbalize a proof. The students in the teaching experiment were provided access through the use of pattern tasks where students could first examine examples before constructing a generalized argument. Some secondary mathematics teachers view proof as a topic of study in geometry or that it should be reserved for the most advanced students (Furinghetti & Morselli, 2011; Knuth, 2002a, 2002b; Kotelawala, 2009), but Lannin shared how proof is possible outside geometry as he worked from what sixth-grade students produced toward communally constructing a valid argument.

Moving beyond beliefs, teachers also want to know how to productively engage students in reasoning and proving. An important part of engagement is selecting a task (NCTM, 2014; Smith & Stein, 2011). Following the reasoning-and-proving framework (i.e., Stylianides, 2008), students are afforded access to proof when they are familiar with reasoning activities similar to those that mathematicians follow. Since students may not know to generate examples on their own in order to identify patterns, the written task can promote reasoning activities explicitly before prompting students for a proof. Karunakaran and colleagues (Karunakaran, Freeburn, Konuk, & Arbaugh, 2014) taught a course to preservice secondary teachers (PSTs) while implementing the Cases of Reasoning and Proof in Secondary Mathematics1 (CORP) materials (Smith, Boyle, Arbaugh, Steele, & Stylianides, 2014) to develop teachers’ (practicing and/or preservice) knowledge and ability to implement reasoning-and-proving tasks. In their study, Karunakaran et al. determined that they were able to improve PSTs’ ability to assess and construct proofs and noted, similar to Lannin (2005), that the tasks used supported PSTs’ access to developing arguments.

As students work toward developing a proof, the classroom community should collaborate to assess proposed arguments to determine what is valid. A. Stylianides

---

1 A set of curricular materials developed under the auspices of the National Science Foundation-funded CORP (Cases of Reasoning and Proof in Secondary Mathematics) Project, grant DRK–12 #0732798, directed by Margaret (Peg) Smith and Fran Arbaugh.
(2007a) developed general criteria of proof in school mathematics:

**Proof is a mathematical argument, a connected sequence of assertions for or against a mathematical claim, with the following characteristics:**

1. It uses statements accepted by the classroom community (*set of accepted statements*) that are true and available without further justification;

2. It employs forms of reasoning (*modes of argumentation*) that are valid and known to, or within the conceptual reach of, the classroom community; and

3. It is communicated with forms of expression (*modes of argument representation*) that are appropriate and known to, or within the conceptual reach of, the classroom community. (p. 291)

The idea is that teachers would use A. Stylianides’s characteristics to communally develop classroom criteria of proof as a way to critique arguments. Moreover, the aim is “to achieve a defensible balance between two (often competing) considerations: *mathematics as a discipline and students as mathematical learners*” (Stylianides, 2007b, p. 294). Teachers assume the role as a representative of the mathematics community to support students with developing criteria and apply the accepted criteria to critique presented arguments.

While a goal would be to have students construct proofs, a learning outcome is for teachers to understand that proofs are accessible for all students across all content and to recognize how to support students’ engagement in constructing and critiquing arguments.

Healy and Hoyles (2000) studied high school students’ ability to construct and assess arguments with respect to proof. The findings include accepting and constructing empirical arguments as proof, but the authors also found that successful provers were those who relied on everyday language as opposed to trying to compose a formal solution consisting of algebraic symbols. An interesting take-away is that the students accepted the algebraic nonproof arguments as proof since they thought those were the ones their teachers would give the highest grade. Therefore, students intuitively understand the value of general arguments that promote understanding, which is what should be the role of proof in classrooms (Hanna, 1995; Hersh, 1993); thus, teachers need to learn to promote a sense-making process of proof where validation goes beyond examples.

### Designing an Instructional Sequence to Reach Learning Outcomes

G. Stylianides and A. Stylianides (2009) designed and implemented an instructional sequence to foster teachers’ understanding that while examples are helpful with finding patterns, arguments based solely on examples fall short of proof. Their instructional sequence was revised over a 4-year period and implemented in an undergraduate mathematics course for elementary teachers. Practically, the instructional sequence included a four-part design across three different mathematics tasks and a reflection after the first three. The fourth part involved the participants in revisiting the initial task to develop a viable argument. Similar to Lannin (2005), G. Stylianides and A. Stylianides found that with scaffolding, the class community was able to construct a proof.

A main theoretical component of the G. Stylianides and A. Stylianides (2009) design was based on cognitive conflict to have preservice teachers experience the limitation of empirical arguments toward a more secure form of validation. Their assumption was that while solving the first task, the preservice elementary teachers would become satisfied with a generalization based on a few examples. The next two tasks support students with realizing the limitations of empirical arguments. The learning outcome is that empirical arguments are insufficient, which is the cognitive conflict, since the assumption is that learners start with assuming empirical arguments are proof.

Similar to G. Stylianides and A. Stylianides (2009), we designed an instructional sequence, but focused on transforming perceptions of proof away from an authoritative toward a communal perspective. Theoretically, our design draws on the situated learning perspective, where knowledge is shared and negotiated communally (Lave & Wenger, 1991). The design assumes that mathematics and mathematics education majors would produce a range of empirical to deductive nonproof and proof arguments depending on their prior experiences. We surmised that students would collectively accept general arguments that are understandable and convincing, and as a class community would dispute those that were based on examples and nonproof deductive arguments along with the instructors’ guidance. The participation in discussions would act on and alter participants’ current thinking about what is an acceptable proof. Given this theoretical approach, our practical design included four learning activities (the four diamonds), as shown in Figure 3.
Each of our four activities was designed to transform perspectives away from authoritative toward a communal perspective of learning. First, we asked participants to prove a non-geometric contextual situation. We wanted teachers to learn that for students to learn to construct proofs then they need to be doing the proving (Karunakaran et al., 2014; Lannin, 2005; Smith et al., 2014; G. Stylianides and A. Stylianides, 2009). The first diamond in Figure 3 is positioned between both the practices and the beliefs rectangles, as it aligns with and addresses both beliefs and engagement. Second, for participants to learn what arguments are considered valid, then they need to do the assessing (e.g., Healy & Hoyles, 2000; Karunakaran et al., 2014; Smith et al., 2014). Third, as the participants evaluated each presented argument, they included their rationale for why each argument was or was not valid. The rationales from analyzing each specific argument are generalized within small groups to form criteria for proof. The small group criteria are synthesized into one list of communal criteria. The idea here is that the participants will begin to reflect on their originally produced argument in light of the criteria. Finally, they use the criteria to explicitly assess their original argument and revise it to construct a proof. It is important to note that these activities relate to those designed in the CORP² project (Smith et al., 2014). Based on this conceptual description, we share our specific implementation in the next section.

**Methods**

The participants (N = 58) in this study spanned four separate courses across four universities and included both undergraduate and graduate students. The authors designed the instructional sequence and were the instructors of the four courses. Two of the courses (taught by Authors Boyle and Ko) were secondary mathematics methods classes, and the other two (taught by Authors Bleiler and Yee) were mathematics content courses. One of the mathematics content courses was specifically for students planning on becoming teachers (taught by Author Yee), and the other mathematics content course was an introduction to proof course for mathematics majors (taught by Author Bleiler). Therefore, the participants in this study were either pursuing undergraduate degrees in mathematics (N = 24) or working toward becoming secondary mathematics teachers at the time of this study.

---

2 The first author was supported by the CORP grant and participated in the development of the CORP materials, so his contributions to the design of this instructional sequence were strongly influenced by his work on the project.
(N = 34). However, since the students across the four courses were not all in a teacher preparation program at the time of this study, we will refer to them more generally as participants.

The four-part instructional sequence designed for this study was implemented across three separate time segments: Before Class, During Class, and After Class, as depicted in Figure 4. It was implemented early in the semester across the four classes. Prior mathematical experiences varied among the participants, but they had all completed at least the first two calculus courses and had completed an introduction to proof course except those in an introductory proof course (Author Bleiler). Most of the participants had not experienced an inquiry-style mathematics class (i.e., Smith, 2006), and some participants in two of the courses (Author Yee [content] and Author Ko [methods]) were initially opposed to working in a format of communal engagement.

Before Class

The Before Class component was a take-home assignment distributed 3 to 7 days before the During Class activity. Participants were instructed to solve the Sticky Gum problem (Fendel, Resek, Alper, & Fraser, 1996) (see Figure 5). In addition to solving the task, they answered the following questions that aimed to promote reflection of their prior experiences about what counts as proof and how they had previously learned about proof:

1. What do you believe are some characteristics of arguments that count as proof?
2. What do you believe are some characteristics of arguments that do NOT count as proof?
3. Based on your past learning experience with mathematical proof (either high school or college), how did you learn about what makes a good mathematical proof? Be specific in your response.

Participants were instructed to submit their responses electronically at least 2 days prior to the During Class activity. Each instructor then purposefully selected five solutions to the Sticky Gum problem from the submitted work. These five solutions would be presented and analyzed in the During Class activity. More specifically, we wanted the participants to evaluate the five arguments and then, based on those evaluations, develop a communally accepted set of characteristics to serve as the class criteria of proof.

Each instructor selected the five arguments that would highlight the diversity of the participants’ responses in their class (see appendixes A–D for all five selected responses from each class). Moreover, instructors selected arguments that they believed had the potential to provoke debate during the class discussion, such as about accepting empirical arguments as proof and the tendency to evaluate arguments based on form. For example, Argument 3 from Author Yee (see Appendix C) is an example of an argument that “looks” like a proof in form, but it lacks mathematical sense and fails to address the problem statement. Argument 2 in Author Bleiler’s class (see Appendix B) is an empirical argument that combines algebraic symbols and seems to be attempting to follow a structure of proof by mathematical induction. Additionally, each instructor included arguments that spanned a

![Figure 4. Implemented instructional sequence to change perspective of proof.](image-url)
Transforming Perceptions of Proof

During Class

At the beginning of the During Class activity each instructor asked the entire class, “Based on your past learning experience with mathematical proof (either high school or college), how did you learn about what makes a good mathematical proof?” Participants in each class were then asked to share some of their responses from this question. After this discussion, they worked individually to decide whether each of the five arguments (Appendixes A–D) was or was not a proof and to provide a rationale for their decision. Then participants were placed in groups of three or four to discuss their assessments and rationales. They were instructed to go beyond labeling “proof” or “not proof” and develop a rationale for each decision. Finally, they drew on the insights from their discussions to create three to seven characteristics that they believed were important for constructing a proof. During these small-group discussions, the instructor asked questions within the small groups about how they assessed each argument and/or listened while the individuals in the small groups shared and discussed their rationales. These interactions among the groups allowed for the participants to verbally articulate their positions and supported the instructor with learning how the participants within each group and across the groups were thinking about the validity of each argument.

A table is included in each appendix (A–D) to show how all of the participants in each course individually evaluated the five arguments before their small-group discussions. A “1” in a cell indicates that the participant identified the argument as proof, and a “0” indicates that the participant labeled the argument as nonproof. The top row of each table in all four appendixes includes the participant that created the argument. For example, in Appendix A, three of the participants initially labeled their own argument as proof.

To launch the whole-class discussion, each group posted its list of characteristics of proof, and the instructor asked the participants to compare and contrast their list with the others. Common themes were shared across the groups while the instructor facilitated the conversation to create a “common class list of characteristics for good proof writing.” We then used the “common list” to evaluate two arguments as a whole class where most students thought the argument counted as proof and some assessed it as a nonproof argument. Using the criteria to evaluate an argument served two purposes: (1) to determine if the characteristic is sufficient to determine the truth of an argument.
argument and (2) to support the entire community with understanding about what the listed characteristics mean with respect to assessing an argument. Therefore, the goal was not to reach 100% agreement, but we believe that having participants engage in the assessment process develops their sense of having a voice within the community and develops their ownership of the criteria.

As instructors, we agreed to limit the number of criteria to five to fit the time limitations of the classrooms and to establish an initial set of agreed upon characteristics. While the criteria that emerged in each class were different (see Figure 6), we aimed to engage the participants in active reflection on what they believe should count as proof based on their classroom community, instead of reaching consensus on the criteria across the four classes (cf. Stylianides, 2007a). The characteristics are arranged vertically to align with each class and horizontally to highlight commonalities across the classes (see Figure 6).

We were careful not to create an environment where any characteristic was acceptable. Our aim was to balance generating communally accepted characteristics while also attending to what is accepted within the larger mathematics community (A. Stylianides, 2007a, 2007b). As instructors, we drew on our expertise along with the participants’ experiences to negotiate individual characteristics to build a shared understanding (i.e., A. Stylianides, 2007a, 2007b). For instance, when Author Boyle’s students offered a characteristic, he wrote it and then asked other participants to share their thinking. Since he questioned their thinking during small-group discussions, he could strategically select particular perspectives to be shared across the class in addition to allowing volunteers to share. In some cases, the original wording was modified before adding a new characteristic to the list. For example, one participant suggested the characteristic of a “clearly defined domain,” as shown in column 1 of Figure 6. Later in the discussion, another participant added the need to “state definitions.”

<table>
<thead>
<tr>
<th>Author Boyle’s Methods Class</th>
<th>Author Bleiler’s Mathematics Class</th>
<th>Author Yee’s Mathematics Class</th>
<th>Author Ko’s Methods Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical Progression. No mathematical errors.</td>
<td>Chain of evidence (clear, step-by-step logical flow). Generalization (Does proof hold true for all cases and answer “why?”).</td>
<td>Organization – Structure of argument. Convincing-Exhaustive, not ambiguous, audience appropriate.</td>
<td>Proof needs to follow a logical order. Proof should be always true for any case and not be verified by specific examples.</td>
</tr>
<tr>
<td>Clearly stated conjecture / hypothesis statement.</td>
<td>Valid/true/correct. Correctness – Supporting evidence, overall structure. (proof built from definitions, theorems, etc.)</td>
<td>Proof needs to include clear explanations that are concise. Proof includes clear statement of what you are trying to prove.</td>
<td></td>
</tr>
<tr>
<td>State a conclusion that follows from the argument.</td>
<td></td>
<td></td>
<td>Argument clear for audiences to follow based on their community</td>
</tr>
</tbody>
</table>
The instructor returned to the former participant’s characteristic and added “define definitions” to merge the two ideas into one category. That is, some of the characteristics were pieced together from multiple participants, and the instructor organized the criteria that was discussed and accepted within the community.

**After Class**

After developing the classroom-based criteria of proof, all participants were asked to complete the *After Class* activity using their class-constructed criteria of proof (see Figure 6). The *After Class* activity was another out-of-class activity where the participants would evaluate their original arguments to the Sticky Gum problem based on their class criteria of proof, revise their arguments, and reflect on their experiences throughout this instructional sequence.

The focus of this paper is on the participant responses to questions 7 and 8 in the *After Class* activity (see Figure 7). A total of 52 of the 58 participants responded to the two reflection questions, and the written responses were analyzed following the principles and techniques of grounded theory (Glaser & Strauss, 1967; Strauss & Corbin; 1990) to develop themes. We chose to ask open-ended reflection questions (questions 7 and 8 in Figure 7), since we did not want to assume past experiences, and we wanted to learn if new themes (related to our hypothetical learning outcomes in Figure 2) might arise that we did not

---

**Postactivity Assignment: Applying class criteria of proof to your argument**

**Name: ________________________________**

*Please complete this assignment by answering the four questions below with a full description. All responses must be submitted digitally. You may scan and email your response.*

**Question 5.** Copy our class-developed criteria for a compelling/convincing argument in the first column, and then complete the remainder of the table based on your original written argument for the Sticky Gum problem.

<table>
<thead>
<tr>
<th>Class-developed criteria for proof writing</th>
<th>On a scale of 0–5 (0 = not at all, 5 = completely), how well does your argument meet each criterion?</th>
<th>Explain your rationale for your numeric rating by making explicit reference to your original argument.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Question 6.** Revise your original argument based on your response to Question 4 so that your new argument would count as a proof based on our class-developed rubric.

**Question 7.** How has this engagement with the Sticky Gum problem and the classroom activity impacted your understanding of proof?

**Question 8.** Describe the similarities and differences between your past learning experiences with mathematical proof and the series of activities we have done with the Sticky Gum problem.
anticipate. The first author listed the themes and revised them while coding the pairs of responses for each of the 52 participants. After narrowing the number of themes to nine across the two research questions, the second author and first author discussed four participant responses after coding them individually. A “1” was placed in a cell that matched a theme with a response and “0” was listed in a cell where there was no match. We discussed our codes until we reached agreement; then Author Boyle and Author Bleiler independently coded 21 participant responses. During our discussion of these 21 pairs of responses, we decided to eliminate a code and change the wording of two others. This resulted in a difference of two or fewer participant-comments for each of the remaining eight themes. The first author revisited each discrepancy and made a final decision, then coded the remaining 27 pairs of responses. In the next section, we share the most common themes along with samples of the participants’ written responses to answer the two research questions.

Results

The results provide insight into how the participants viewed their experience throughout our instructional sequence compared to their previous experiences with proof, and the ways participants perceived that the activities in the instructional sequence changed their perceptions of proof. Representative responses from the reflection questions (questions 7 and 8 in Figure 6) across the four courses are included to highlight how the participants explained their thinking. It is important to state that participants’ responses were coded under a particular theme only when they explicitly wrote about that theme. Therefore, it is possible that more of the participants could find some of the same themes to have an impact on their learning, but they were not specifically asked to comment on them, given the general nature of the questions. Additionally, a pseudonym is listed after each reflection and the instructor of the participant is in parentheses. Portions of some responses are italicized. This is done to highlight the connection with the theme. We include the full response or a majority of each response to share how most of the responses span several themes. We conclude this section by sharing four cases to compare a participant’s written argument against his or her reflection responses.

RQ1: In what ways do participants perceive this instructional sequence as different in relation to their prior experiences with mathematical proof?

Table 1 includes the five most common themes for the 52 participants related to our first research question. We explain each of the five themes and provide sample responses. The first three themes align with the tasks in our instructional sequence. The fourth theme relates to how participants engaged in the tasks. The fifth theme aligns with how some participants have previously engaged in proof. The total number of responses exceeds 52 since many of the participants referenced more than one perceived difference.

Discussing what is needed for an argument to count as proof. Most participants (73%) had never worked from implicit or explicit criteria of proof, and they explained it was helpful for them to generate characteristics of what counted as proof. They identified the development of common criteria as a supportive activity that helped them to construct a holistic view of proof. Beyond just identifying the usefulness of criteria in general, the three participants below spoke about specific characteristics such as clarity and generalizing beyond a set of examples. Gina also adds that she now has a greater understanding of what the characteristics mean. The following responses are representative of how most participants reflected on the development of criteria as being different and important to them:

I have never seen a rubric for a proof. Creating and thinking about a rubric helped me make sure that my proof was of quality for someone to read and understand. This rubric helped guide my thinking in creating and evaluating proofs, but

<table>
<thead>
<tr>
<th>Theme 1</th>
<th>Theme 2</th>
<th>Theme 3</th>
<th>Theme 4</th>
<th>Theme 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>38/52 (73%)</td>
<td>18/52 (35%)</td>
<td>15/52 (28%)</td>
<td>18/52 (35%)</td>
<td>20/52 (38%)</td>
</tr>
</tbody>
</table>
it did not set complete restrictions on the exact format of the proof. Brandon (Author Ko)

Now I have a better understanding of what the proof, regardless of community, needs to entail. I need to be sure that everyone will be able to follow the explanations clearly, thus they need to be clear and concise. They also need to be logical. Even though I may provide all the explanations and evidence in the proof, they need to be in a logical order so that the audience can follow systematically. Additionally, the proof needs to be generalized. My past experience and this experience are similar in that I realized they must have a conjecture, explanations, and evidence that prove the conjecture. However, I now have a better understanding of what exactly this means. Gina (Author Ko)

I feel that this activity gave me the opportunity to learn the basics of proof. I feel that in the past it wasn’t explained to me that proof writing is not specific. In a sense it gave me an abstract way to write a proof in a way that made it concrete. I really liked the idea of comparing others work so that we could determine the criteria for a proof. I liked that we came up with a list of the criteria for future references. Carmelo (Author Boyle)

Conducting and/or revising an argument. About one third of the participants (35%) indicated that constructing an argument on their own prior to being provided sample arguments and/or being asked to revise their original argument based on the class criteria was different. Many participants mentioned that proof now seems much more accessible to them as learners than it has been in the past. Most important, they appreciated the opportunity to learn from their mistakes as they revised their original argument against the classroom criteria. The following are representative of the participants’ responses:

First, I understand what proof means more clearly now. When rewriting a proof, I tried to incorporate or use what I learned during the class discussion. I tried to generalize it more because I learned that generalizing was my weakest part in my previous proof. This engagement is helping me to understand the problem. Most importantly, I understand that trying and struggling are very important processes in order to fully understand the task. Without trying to solve the task alone, before interruption of teacher, is important for me to find out what my weaknesses and strengths are. Mona (Author Boyle)

In the past, we only worked with examples and then the teacher telling us to do problems when we have been lectured on the right way to do them. In this case, we found our errors basically by ourselves with the teacher prompting and developed our own understanding of what proofs are and now have a better understanding on how to write proofs. We know why we do the things that we do so that it will be a clear concise proof. We saw that if you don’t write proofs well, they either don’t make sense or are very hard to read. Carla (Author Ko)

I thought it was very helpful to go back and revise my proof. Something different that I experienced this time was being able to clearly map out the structure of proof before going back to revise it. This improves the proof tremendously. Stacey (Author Yee)

Analyzing and discussing the validity of arguments. Twenty-eight percent of the participants identified reviewing their classmates’ arguments as new and very supportive to their learning. Through evaluating and discussing the validity of their peers’ work, participants realized that they tended to think differently from one another. The variety of the instructor-selected arguments helped them to identify characteristics that were similar to and different from their solution method and supported them with improving their ability to distinguish between those that qualify and do not qualify as proof. Below is a sample of what was written related to this theme:

It really made me realize how differently this problem can be viewed based on the 5 examples given in class. The same problem can be proved in numerous ways, depending on the reader’s interpretation of the questions. Also, there is no right and wrong way to do it, but rather many ways where some may be more sufficient and effective than others. Jacqui (Author Yee)

The Sticky Gum Problem and the activities that went along with it really taught me the importance of clarity and generalization in a proof. Looking at proofs written by other students through a critical lens really has helped me understand what works and what does not work in a proof. One of the most common mistakes in the Sticky Gum proofs was the use of examples to verify that a general rule was correct. Brandon (Author Ko)

In the past I would just jump to an applicable solution. This often led to an incomplete
knowledge of the proof. Diving into writing and dissecting these proofs has given me a deeper understanding and a tool set to go back to older proofs. Gabriel (Author Bleiler)

Working collaboratively in class. Thirty-five percent of the participants explained that being afforded the opportunity to discuss their thinking with their peers was unusual from prior proof experiences. Furthermore, it helped them discuss the shortcomings of their solution path and negotiate their thinking about what counts as proof. Specifically, they found it empowering to be part of the process with both assessing arguments and generating characteristics for the class criteria. Sharing their thinking about proof with their peers was new and powerful, and this is what a sample of them wrote:

[The activity] has given me time to struggle with what makes a good proof or not by myself and with others. Listening to others has really helped me to see where my thinking and generalizations tend to come from. Danielle (Author Bleiler)

Some differences were the way we went about discussing what makes a proof. Normally, as a student, I am lectured as to what makes a good proof from a teacher/professor; in our class we thought individually, then as a small group, and concluded as a whole class. I like the way we discussed proofs, because then I thought that I was actually making up the rules, rather than being told what to write or expect. Naomi (Author Ko)

I have never had the opportunity to compare other students’ proofs of a problem with my own as a class. I have never discussed what others believe constitutes a valid proof as a class and heard the explanations for why these specific criteria are what others consider important for proof writing. This experience let me see some of my proof writing weakness by allowing others to express their doubts about certain aspects of the proofs our classmates wrote. Antonella (Author Yee)

Previously asked to memorize and reproduce the textbook or instructor proof. More than one third (38%) of the participants shared that their previous experiences focused on reproducing textbook or instructor produced proofs. In other words, proof seemed difficult or confusing since they were never provided an opportunity to engage in the proving process. With the Sticky Gum sequence, the participants explained that developing an argument themselves and learning the limitations of their arguments enhanced their understandings of proof in general. Here is a sample of participants’ thoughts:

The activities we did helped me to feel more comfortable with writing proofs. . . . When proofs were first introduced to me, it was in geometry, and my teacher did most of the writing and we watched. Proofs didn’t come up again until last fall in Calculus III, and again, my teacher did most of the writing. . . . I simply just familiarized myself with different common manipulations for writing a convincing argument. Having to come up with/critique proofs gave me a way of trouble shooting what they are and how to write them. Through learning from mistakes and trial and error I was able to get a better understanding. Calvin (Author Bleiler)

Before class, I had viewed proof writing as a purely academic endeavor that had little or no real world application. Not to sound bitter, but in previous classes, the process of proof seemed like a large amount of academic back-patting and teachers saying, “Hey, look at me, I can do this really complicated piece of math on the board.” Peter (Author Bleiler)

In my past, learning proofs has been very difficult. We were basically shown examples with no guidance as to what should be included in a proof. These activities have helped to demonstrate the key concepts that should be included in proofs. As I look back, all of the key things we described to make up a good proof were included in the proofs my teachers did but weren’t explained. Teachers simply showed us different proofs and then expected us to write some on our own. I think I would have had a much easier time with proofs had I been taught what should be included to make a proof valid. Jessica (Author Ko)

RQ2: How (if at all) do participants perceive that the activities in this instructional sequence changed their perceptions of proof?

Forty-seven of the 52 participants across all four classes identified at least one activity from the instructional sequence as supportive. In this section, we share three themes that address new insights and transformed perspectives of proof (see Table 2). The first theme relates to the participants gaining a clearer understanding about what is needed for an argument to count as proof or how to construct a proof themselves. The second theme
addresses how proofs should be assessed so that the entire community is a participant in the sense making and assessment process. The third theme highlights a previous misconception, namely, an overreliance on format. Many of the participant responses were coded into more than one theme. After sharing a few quotes for each of these three themes, we select one participant from each that had his or her argument assessed to learn how they might be thinking about their initial argument in relation to their reflection responses.

More clear about what counts as proof. Most participants (85%) explained that they never really understood where proofs came from or what was required for an argument to count as proof. They shared that the instructional sequence helped them to better understand what is needed for an argument to count as proof. Some explained that some of the ideas they thought were important were made more explicit and secure for them. Others wrote about how the instructional experience supported them with beginning to learn what counts. While many participants still believe they need more experience, they feel they have a better understanding with how to start and what is expected. Sample responses are given below:

**Before this assignment, I had no clue what made a good proof nor how to write one by myself. Previous math classes only stated the proof and said how it was true. This activity stated out point by point how to make a proof valid and understood. Natasha (Author Bleiler)**

**I feel like as a class now we have a better grasp on what we all believe is a good proof. The activity has broken down proof writing so that it is not as formal. We state what we want to show and then show it. We have guides also to help examine how we can strengthen our proofs off of other criteria and not just what we were used too. Michael (Author Ko)**

**This engagement along with the discussion of proofs has helped me understand proofs a little more. I know how different people will have a different take on what makes a good proof and how to prove it. I also know that most of us only have a basic understanding of proof, and will continue to learn what makes a good proof as well as the process. Naomi (Author Yee)**

Provided with an opportunity to see proof as socially negotiated. Just over half of the participants (53%) identified the social interactions with their peers as supportive to reshaping their thinking. Analyzing their peers’ solutions supported them in realizing that not all viable arguments to this problem need to be done the same way. The social interactions among the participants helped them to begin to gain ownership of proof as something that must adhere to agreed-upon criteria as opposed to just satisfying an instructor. Overall, they seem to now view proofs as arguments that should make sense to them and their community while also attending to the agreed-upon criteria. Below are representative responses:

**This engagement with the Sticky Gum Problem and classroom activity has impacted my understanding of proof by giving me a more solid foundation of what not only a teacher sees as a valid proof, but what my peers see as a valid proof. I was able to compare my work that I created on my own with work from my peers for the very same problem. It helped me realize that initially not everyone goes about proving a problem the same way. I think after having the classroom discussion I understand different aspects of a proof, such as organization and clarity a bit more than I did before. Antonella (Author Yee)**

**This engagement has given me a sort of “foundation.” Knowing that these are the criteria that every proof should meet, I can now check my proofs to make sure they are meeting all the qualifications for a good proof. With the classroom activity I can see how other people are thinking when doing proofs. It’s not just the teacher’s opinion and mine but now it is a community of provers. Hilary (Author Bleiler)**

**In Calculus III we were trying to manipulate an equation to somehow find “delta” to make**
Expanded thinking beyond a particular format. About half of the participants (52%) shared that they previously believed that proof needed to follow a particular format, such as using a certain proof method, or that it needed to include certain symbols or notation. Through participating in this sequence, they learned that the content of what is written matters more than the format. In other words, they realized that learning to prove goes beyond simply employing specific proof techniques and including mathematical symbols. A few specific thoughts are shared below:

"I was under the impression, previously that proofs were subject to “standard” types of proofs. Implying that a type of proof was to be chosen and a strict format to be followed. In this activity there is a format but it is much more flexible than I initially thought, requiring key elements instead of outlines." Henry (Author Ko)

"I feel that this activity gave me the opportunity to learn the basics of proof. I feel that in the past it wasn’t explained to me that proof writing is not specific. In a sense it gave me an abstract way to write a proof in a way that made it concrete. I really liked the idea of comparing others works so that we could determine the criteria for a proof." Carmelo (Author Boyle)

Comparing original arguments with reflection responses: Four case examples. In this section, we provide four case examples, one from each of the four courses. With these cases we intend to illustrate how participants’ reflective comments relate to their initial arguments. By looking more closely at a participant’s initial argument, we can better understand the context of their reflective comments. We selected these cases examples from participants who initially identified their invalid argument as proof (at the beginning of the During Class activity). In addition, we selected the case examples from participants whose responses provided us with insights into how they were making connections to their original arguments.

In Appendix A, Ji-min (Author Boyle, argument 1) individually labeled her argument as proof, although none of her peers believed the argument was valid. After the class discussion she shared thoughts that caused her to change her thinking about the validity. She wrote, “A proof has to hold true for every possible situation. . . . Re-writing my proof I was reminded of how much thought and effort goes into a proof.” Prior to this instructional sequence, she accepted generalizations based on a set of examples as proof. Her original argument was empirical, and after the instructional sequence, she seemed to be clear that her original argument was not true for all cases, which was a listed characteristic in this class.

In Appendix B, the table shows that Keyshawn (Author Bleiler, argument 2) labeled his argument as proof. About 65% of Keyshawn’s peers in his class also labeled his solution valid, even though it included misunderstandings about both proof by mathematical induction and the use of variables. For example, he wrote that the \( x + 1 \) case holds true, but he only checked a specific case (3). After the instructional sequence, he wrote, “Proofs can mean several things; there isn’t one standard for proof writing.” This comment suggests that he changed his perspective away from believing a proof needs to follow a specific format. In the past he only followed what his instructors produced without trying to make sense of proofs. He added, “Past proofs have been much more teacher-led rather than this one which was almost all class-led in group discussions.” Thus, Keyshawn now understands that he needs to write arguments that are acceptable to the class community. While it is not completely evident from what he wrote, it seems as though he realized from the class discussion that his original argument was not a proof.

In Appendix C Andy (Author Yee, argument 3) and about 56% of his classmates labeled his argument as proof. While the argument does not make sense mathematically, he included mathematical symbols and a format that includes multiple cases. After the instructional sequence, Andy wrote, “This activity has given me more of an
understanding on what is an actual proof. As a proof writer I have to be careful to write proofs in a way that is understandable not only to me, but also to whoever is going to read it.” He also added, “Some differences about the activity would be to have more of an open mind about how to do a proof.” Andy does seem to realize that his original argument required revisions, as he wrote that it needs to be understandable to others. Additionally, the comment about being more open-minded seems to suggest that he, like Keyshawn, realizes that proofs are not confined to a specific format. Therefore, Andy may have realized the limitations of trying to make an argument proof like (i.e., using symbols and following a particular structure), and through engagement in the instructional sequence, he has a new view about what is acceptable and should be understandable to others.

Finally, in Appendix D, Brandon (Author Ko, argument 1) and about 30% of his peers labeled his argument as proof. Brandon made a generalization and attempted to explain why his formula works in every possible case, but as written it is invalid. After the instructional sequence, Brandon wrote, “The Sticky Gum Problem and the activities that went along with it really taught me the importance of clarity and generalization in a proof.” Here, like Ji-min, he identifies generalization as a limitation, and this may mean that he realized that his argument fell short since it did not include a generalized argument. He also added, “One of the most common mistakes in the Sticky Gum proofs was the use of examples to verify that a general rule was correct. Although it is good to check your generalization with examples, a proof should be able to stand alone without them.” These comments suggest that Brandon realized that his original argument was not a proof.

These four cases, one from each course, provide additional support that participants shared not only that their perspective of proof changed, but that their new perspective was explicitly aligned with how they originally constructed their argument. For example, Brandon and Ji-min both constructed empirical arguments and shared after the instructional sequence that they now understand proofs need to be true for all cases. In addition, Andy and Keyshawn tried to fit their arguments into some type of deductive form without attending to mathematical meaning. After engaging in the instructional sequence, they both wrote about how a strict format is less important than they had originally thought. All four participants from each of the four courses shared specific comments related to the arguments they originally produced, which suggests that they each have changed their perspectives about how to construct a proof that was different prior to the instructional sequence.

Discussion

Reasoning and proving should be central activities throughout the K–12 curricula, since it has the potential to develop a deeper understanding of mathematics. However, in order for this to become reality, teachers need opportunities to learn reasoning and proving to support their change away from the authoritative perspective, as it contributes to limiting perceptions and misunderstandings about the nature of proof. We shared an instructional sequence that we implemented with mathematics and mathematics education students. The mathematics task promoted access to proof since it explicitly calls students to examine cases leading to the development of a wide variety of arguments based on their prior experience. As expected, none of the participants explained that they did not know how to start the task, but many realized what they originally produced required revision to count as proof. Through active engagement with constructing an argument, analyzing a set of five arguments, and developing a communal proof criteria, many participants were able to identify their own prior misunderstandings and share how the learning opportunity supported them in gaining a better understanding about what is needed for an argument to count as proof. The 24 responses in the results section across the four different classes provide examples of the impact of the instructional sequence.

The reflections suggest that the participants experienced movement away from the authoritative perspective of proof and that they are embracing a communal perspective as learners, as shown in Table 3. It is encouraging to read how some participants came to dismiss previous beliefs, such as proofs needing to follow a particular format. Some also shared that assessing their peers’ arguments helped them to rethink their approach to the problem and specifically realize why empirical arguments are not proof. Previous research reported that teachers hold narrow and limiting views of proof (e.g., Karunakaran et al., 2014; Knuth, 2002b; Steele & Rogers, 2012), and these perceptions have an impact on their instructional practices (Bieda, 2010; Furinghetti & Morselli, 2011). We agree with Harel and colleagues (Harel & Sowder, 1998; Harel & Rabin, 2010) that when secondary or undergraduate students experience proof in authoritative settings, they develop limiting views of proof and these views stay with them even after they become teachers. We believe the research-based design of our instructional sequence jumpstarted a transformation from the authoritative toward a communal perspective of proof. The participant reflections highlight some of the key distinctions between authoritative and communal engagement with proof, which we sought to capture in Table 3.
We view authoritative engagement and communal engagement as existing along a spectrum, rather than being two isolated options. Believing that students should be afforded proving opportunities is a productive first step, but productive engagement and assessment routines are required to promote student learning. Productive engagement requires an accessible task that allows for multiple solution paths (NCTM, 2014; Smith & Stein, 2011). Time is also required for students to analyze strategically selected arguments. Students should be encouraged to share their current thinking, as opposed to the “correct solution,” for the community to assess and provide feedback. We believe that these negotiated, communal, sense-making engagement experiences afford opportunities for our participants to change their thinking about how they view proof (shown in column 4 of Table 3). Most, if not all, of the participants experienced changes in their perspective of proof from an individual and or memorization activity that must follow a strict (possibly single) format to viewing proof as a collaborative process where proposed arguments are critiqued and revised based upon negotiated criteria. For example, participants stated that they previously experienced proof as passive recipients in settings where their instructor provided the correct proof and that a proof needed to follow a particular structure or format. These prior experiences could be why the participants also thought that proof had to follow a specific format or structure. Our results suggest that transforming engagement with proof can lead to important changes in learners’ perceptions of proof. Just as authoritative learning experiences explain why teachers develop narrow beliefs and in turn repeat a similar authoritative enactment, we hope that this learning experience can cause teachers to change their instructional practices toward a communal perspective of proof.

We are excited by the results of this study, and although the data we have analyzed here suggest participants’ movement in a positive direction, we realize that more work is needed. These participants displayed similar mathematical misunderstandings found in studies with secondary students, such as accepting empirical arguments as proof and believing that the form (e.g., including specific symbols) is more important than making sure others can follow and make sense of your argument. The four cases shared in the results section highlight how the instructional sequence seems to have affected their thinking about the limitations of their original arguments. Even though the instructional sequence supported them with expanding their understanding about what counts as proof, many participants realized that more experience is needed to develop their ability to construct proofs. For instance,

We view authoritative engagement and communal engagement as existing along a spectrum, rather than being two isolated options. Believing that students should be afforded proving opportunities is a productive first step, but productive engagement and assessment routines are required to promote student learning. Productive engagement requires an accessible task that allows for multiple solution paths (NCTM, 2014; Smith & Stein, 2011). Time is also required for students to analyze strategically selected arguments. Students should be encouraged to share their current thinking, as opposed to the “correct solution,” for the community to assess and provide feedback. We believe that these negotiated, communal, sense-making engagement experiences afford opportunities for our participants to change their thinking about how they view proof (shown in column 4 of Table 3). Most, if not all, of the participants experienced changes in their perspective of proof from an individual and or memorization activity that must follow a strict (possibly single) format to viewing proof as a collaborative process where proposed arguments are critiqued and revised based upon negotiated criteria. For example, participants stated that they previously experienced proof as passive recipients in settings where their instructor provided the correct proof and that a proof needed to follow a particular structure or format. These prior experiences could be why the participants also thought that proof had to follow a specific format or structure. Our results suggest that transforming engagement with proof can lead to important changes in learners’ perceptions of proof. Just as authoritative learning experiences explain why teachers develop narrow beliefs and in turn repeat a similar authoritative enactment, we hope that this learning experience can cause teachers to change their instructional practices toward a communal perspective of proof.

We are excited by the results of this study, and although the data we have analyzed here suggest participants’ movement in a positive direction, we realize that more work is needed. These participants displayed similar mathematical misunderstandings found in studies with secondary students, such as accepting empirical arguments as proof and believing that the form (e.g., including specific symbols) is more important than making sure others can follow and make sense of your argument. The four cases shared in the results section highlight how the instructional sequence seems to have affected their thinking about the limitations of their original arguments. Even though the instructional sequence supported them with expanding their understanding about what counts as proof, many participants realized that more experience is needed to develop their ability to construct proofs. For instance,
Recommended Next Steps for Teacher Educators

A strength of this instructional sequence is that it was implemented at four different institutions with participants possessing varying degrees of mathematical experience. Regardless of the class context across the four instructors, the participants shared how the instructional sequence affected their perspective of proof. One of the primary challenges students across varying grade levels face in developing an understanding of proof is the authoritative proof scheme (Harel & Sowder, 1998); instructors often perceive that their students are unable to produce a proof on their own (Harel & Sowder, 2009) causing them to decide to present their own proofs for students to reproduce (Stylianou, Blanton, & Knuth, 2009). This instructional sequence acknowledges this instructor-student tension but embraces the perspective that learning meaningfully requires active student engagement. Although providing learners with opportunities to collaborate with peers and make sense of mathematics takes time, the result shows that participants gain a deeper understanding of the nature of proof and, therefore, likely of the nature of mathematics.

While participants pointed to specific activities such as developing a criterion or working collaboratively, we believe the collection of activities sequenced in the way described is critical to gaining similar results. In other words, we find it advantageous to start with having learners solve a task before developing criteria for proof or evaluating the validity of arguments. While the Sticky Gum problem does not need to be the task used in the sequence, we recommend choosing a task that provides opportunities to look for patterns, generate examples, make a generalization, and construct an argument. These reasoning-and-proving activities (Stylianides, 2008) provide access into the problem, promote different types of solution paths, and allow for misconceptions to surface, such as proofs based on empirical examples. As Lannin (2005) found with middle school students and G. Stylianides and A. Stylianides (2009) with undergraduate elementary majors, others will most likely need support with developing a general argument that connects their formula to the general context of the problem. Having learners evaluate their peers’ argument may seem contentious, but explicitly explaining that their work has the potential to be shared and that their names would be removed can help them feel comfortable. Also, this practice seemed to directly support Ji-min, Keyshawn, Andy, and Brandon (four cases previously shared) as we compared their original solution against their reflection after engaging in the instructional sequence. One possible extension to the sequence might be to spend time reviewing and discussing students’ revised arguments with respect to the criteria.

In moving this work forward, we believe it is important to continue to think critically about how this new knowledge about proof can be parlayed to secondary mathematics teachers supporting their students to construct and critique arguments (CCSSM, 2010). While some participants realized this next step, it was not an explicit part of this instructional sequence. For instance, Francine in Author Ko’s class wrote:

“This particular problem has really opened my eyes to how broad and important proofs can be. Proofs are not only used to prove theorems or equations to be true. I also understand now that proofs can be used from very early on in education rather than only in geometry at the high school level.”

Therefore, some participants did realize our ultimate implicit goal, but more work is needed to support participants after changing their perceptions of proofs to also transfer this instructional perspective into secondary classrooms to develop students’ capabilities in constructing and critiquing arguments. Nonetheless, independent of the generalizability of our qualitative results, the versatility of this activity sequence can allow teacher educators to genuinely connect with teachers and help them view teaching and learning proof as a communally constructed, sense-making mathematical activity.

References


Steele, M. D., & Rodgers, K. C. (2012). The relationship between mathematical knowledge for teaching
Transforming Perceptions of Proof


**Authors**

Justin D. Boyle, Department of Curriculum and Instruction, 230D Graves Hall, University of Alabama, Tuscaloosa, AL 35487; jboyle@bamaed.ua.edu

Sarah K. Bleiler, Department of Mathematical Sciences, 263 Kirksey Old Main (KOM), Middle Tennessee State University, Murfreesboro, TN 37132; sarah.bleiler@mtsu.edu

Sean P. Yee, Department of Mathematics & Instruction and Teacher Education, LeConte College 317K and Wardlaw 257, University of South Carolina, Columbia, SC 29208; yeesp@mailbox.sc.edu

Yi-Yin (Winnie) Ko, Department of Math and Computer Science, Root Hall A-140E, Indiana State University, Terre Haute, Indiana, 47809; winnie.ko@indstate.edu
Appendix A: Author Boyle

Argument 1

We may come up with the following chart using the number data (3).

<table>
<thead>
<tr>
<th></th>
<th>C: # of color</th>
<th>F: # of family size</th>
<th>P: # of penny needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>10</td>
<td>13</td>
</tr>
</tbody>
</table>

F (# of family) means that we have to get the same # of same color gumball. In diagrams, we may notice that we have to repeat \((F-1)\) times of subset and add 1 time at the end to get \(F\)th gumball.

In Subsets (hw), there are all different color gumballs (C: # of colors).

Therefore, we can infer this diagram and number chart to one formula (or equation):

\[ P = C \times (F-1) + 1 \]
Argument 2

Please complete this assignment by answering the three questions below with a full description. All responses must be submitted digitally. You may scan and email your response.

Sticky Gum Problem

Ms. Hernandez came across a gumball machine one day when she was out with her twins. Of course, the twins each wanted a gumball. What’s more, they insisted on being given gumballs of the same color. The gumballs were a penny each, and there would be no way to tell which color would come out next. Ms. Hernandez decides that she will keep putting in pennies until she gets two gumballs that are the same color. She can see that there are only red and white gumballs in the machine. (The following questions are to help guide you towards answering question 1. You should complete all of the bulleted questions, but you only need to email complete responses for Questions 1-3 below. Based on your solution to question 1 below, I may ask you to turn in your bulleted responses in class.)

- Why is three cents the most Ms. Hernandez will have to spend to satisfy her twins?
- The next day, Ms. Hernandez passes a gumball machine with red, white, and blue gumballs. How could Ms. Hernandez satisfy her twins with their need for the same color this time? That is, what is the most Ms. Hernandez might have to spend that day?
- Here comes Mr. Hodges with his triplets past the same gumball machine with red, white, and blue gumballs. Of course, all three of his children want to have the same color gumball. What is the most he might have to spend?

Question 1. Generalize this problem as much as you can. Vary the number of colors. What about different size families? Prove a generalization that always works for any number of children and any number of gumball colors. I encourage you to think about using alternative representations within your proof in order to create your most convincing/compelling argument. R-red, W-white, B-blue, Y-yellow, G-green these are the colors of gumballs

<table>
<thead>
<tr>
<th>RWB/RWB/RWB/RWB/RWB</th>
<th>RWBY/RWBY/RYWR/YBRW/RWYB</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 gum balls, 2 kids, spend $0.04</td>
<td>4 gum balls, 2 kids, spend $0.05</td>
</tr>
<tr>
<td>3 gum balls, 3 kids, spend $0.07</td>
<td>4 gum balls, 3 kids, spend $0.09</td>
</tr>
<tr>
<td>3 gum balls, 4 kids, spend $0.10</td>
<td>4 gum balls, 4 kids, spend $0.13</td>
</tr>
<tr>
<td></td>
<td>4 gum balls, 5 kids, spend $0.17</td>
</tr>
</tbody>
</table>

RWBYG/RGBYW/YBGRW/WBYRG/GBWRY

For every child you add, the amount of money you spend increases by the number of colored gum balls present.
Argument 3

Question 1:
Conjecture: For any number of kids, \( k \) and any number of gumball colors \( c \), where each kid must have the same gumball color the total worst case scenario cost at 1¢ per gumball is \( c(k-1) + 1 \)

Assume the same color is drawn for each kid except one kid because if every kid got the same color it would not be the worst possible case. Then a new streak of color (2nd color) is drawn for every kid, but again it stops one short to match every kid. Then a third color begins to be drawn from the machine and again the number of gumballs of this third color is one less than is needed.

Here let me show you a picture of what I mean.

One less than the total number of kids (\( k-1 \))

Color 1

... …

One less than the total number of kids (\( k-1 \))

Color 2

... …

One less than the total number of kids (\( k-1 \))

Color 3

... …

One less than the total number of kids (\( k-1 \))

Color c

... …

Where \( c \) is the total number of colors and is a natural number

So after every gumball color, \( c \) is drawn \( k-1 \) times then every kid but one has a matching gumball and since there are no new colors in the machine, the very next gumball (+1) will be a match so that every kid has the same color gumball at the worst possible case.

So the conjecture holds true \( c(k-1) + 1 \)
Argument 4

When I increased \# colors and keep \# kids constant, but increase \# kids to 3 from last example, maximum spent went down. The following chart represents this:

<table>
<thead>
<tr>
<th>COLORS</th>
<th>KIDS</th>
<th>Maximum Spent ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

A general equation that I found that worked for \# colors and \# kids to = maximum spent is the following:

\[
\left[\text{(\# of colors)} \times \text{(\# of kids)}\right] - \left[(\# of colors) - 1\right]
\]
Argument 5

If we always consider the longest possible chain of events that gets us to our desired end, and we count each draw as $\frac{1}{10}$, we find the number of cents needed to get $n$ balls of the same color.

$\text{Set } \# \text{ of colors } = C$

$\text{Set } \# \text{ of children } = n$

$\text{Set } \# \text{ of all colors } = C$

$\text{Set } \# \text{ of colors of each color }$

$\text{Interchangeable combinations } = n!$

We know this is the maximum because completing a set of some number of gumballs of the same color any earlier means you will have $n$ gumballs of one color sooner than in the model above, meaning you’ve spent fewer cents than in the model above.

This rule works for $n \geq 1$

<table>
<thead>
<tr>
<th></th>
<th>A1 - Ji-min</th>
<th>A2 – Mona</th>
<th>A3 - instructor</th>
<th>A4 - Linda</th>
<th>A5 – Natalia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carmelo</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Darren</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Linda</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Natalia</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Mona</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ji-min</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>0.167</strong></td>
<td><strong>0</strong></td>
<td><strong>1</strong></td>
<td><strong>0.167</strong></td>
<td><strong>0.833</strong></td>
</tr>
</tbody>
</table>

(Return to page 38)
Appendix B: Author Bleiler

Argument 1

Ms. Hernandez came across a gumball machine one day when she was out with her twins. Of course, the twins each wanted a gumball. What’s more, they insisted on being given gumballs of the same color. The gumballs were a penny each, and there would be no way to tell which color would come out next. Ms. Hernandez decides that she will keep putting in pennies until she gets two gumballs that are the same color. She can see that there are only red and white gumballs in the machine. (The following questions are to help guide you towards answering question 1. You should complete all of the bulleted questions, but you only need to email complete responses for Questions 1-4 below. Based on your solution to Question 1 below, I may ask you to turn in your bulleted responses in class.)

- Why is three cents the most Ms. Hernandez will have to spend to satisfy her twins?
- The next day, Ms. Hernandez passes a gumball machine with red, white, and blue gumballs. How could Ms. Hernandez satisfy her twins with their need for the same color this time? That is, what is the most Ms. Hernandez might have to spend that day?
- Here comes Mr. Hodges with his triplets past the same gumball machine with red, white, and blue gumballs. Of course, all three of his children want to have the same color gumball. What is the most he might have to spend?

**Question 1.** Generalize this problem as much as you can. Vary the number of colors. What about different size families? Prove a generalization that always works for any number of children and any number of gumball colors. I encourage you to think about using alternative representations within your proof in order to create your most convincing/compelling argument.

\[
\begin{align*}
C &= \text{children} \\
\$ &= \text{cost} \\
X &= \text{number of different colored gumballs} \\
\$ &= (C-1) \cdot X + 1
\end{align*}
\]
Argument 2

Ms. Hernandez came across a gumball machine one day when she was out with her twins. Of course, the twins each wanted a gumball. What’s more, they insisted on being given gumballs of the same color. The gumballs were a penny each, and there would be no way to tell which color would come out next. Ms. Hernandez decides that she will keep putting in pennies until she gets two gumballs that are the same color. She can see that there are only red and white gumballs in the machine. (The following questions are to help guide you towards answering question 1. You should complete all of the bulleted questions, but you only need to email complete responses for Questions 1-4 below. Based on your solution to Question 1 below, I may ask you to turn in your bulleted responses in class.)

- Why is three cents the most Ms. Hernandez will have to spend to satisfy her twins?
- The next day, Ms. Hernandez passes a gumball machine with red, white, and blue gumballs. How could Ms. Hernandez satisfy her twins with their need for the same color this time? That is, what is the most Ms. Hernandez might have to spend that day?
- Here comes Mr. Hodges with his triplets past the same gumball machine with red, white, and blue gumballs. Of course, all three of his children want to have the same color gumball. What is the most he might have to spend?

**Question 1.** Generalize this problem as much as you can. Vary the number of colors. What about different size families? Prove a generalization that always works for any number of children and any number of gumball colors. I encourage you to think about using alternative representations within your proof in order to create your most convincing/compelling argument.

\[
\text{# of colors} = n, \quad \text{# of children} = x
\]

\[
n(x-1) + 1 = T(n,x)
\]

- Mrs. Hernandez w/ R, W gumballs: \( n=2, x=2 \)
  \( T = 2(2-1) + 1 = 3 \)

- Mrs. Hernandez w/ R, W, B gumballs: \( n=3, x=2 \)
  \( T = 3(3-1) + 1 = 8 \)

- Mr. Hodges w/ R, W, B gumballs: \( n=3, x=3 \)
  \( T = 3(3-1) + 1 = 7 \)

The expression \( T(n,x) = n(x-1) + 1 \) is valid for the values of \((n=2, x=2)\) and \((n+1=3, x=2)\) as well as \((n+1=3, x+1=3)\), so we can assume the expression is valid for all positive integers of \(n\) and \(x\).
Argument 3

Ms. Hernandez came across a gumball machine one day when she was out with her twins. Of course, the twins each wanted a gumball. What’s more, they insisted on being given gumballs of the same color. The gumballs were a penny each, and there would be no way to tell which color would come out next.

Ms. Hernandez decides that she will keep putting in pennies until she gets two gumballs that are the same color. She can see that there are only red and white gumballs in the machine. (The following questions are to help guide you towards answering question 1. You should complete all of the bulleted questions, but you only need to email complete responses for Questions 1-4 below. Based on your solution to Question 1 below, I may ask you to turn in your bulleted responses in class.)

- Why is three cents the most Ms. Hernandez will have to spend to satisfy her twins?
- The next day, Ms. Hernandez passes a gumball machine with red, white, and blue gumballs. How could Ms. Hernandez satisfy her twins with their need for the same color this time? That is, what is the most Ms. Hernandez might have to spend that day?
- Here comes Mr. Hodges with his triplets past the same gumball machine with red, white, and blue gumballs. Of course, all three of his children want to have the same color gumball. What is the most he might have to spend?

Question 1. Generalize this problem as much as you can. Vary the number of colors. What about different size families? Prove a generalization that always works for any number of children and any number of gumball colors. I encourage you to think about using alternative representations within your proof in order to create your most convincing/compelling argument.

Let $x$ = # of colors
$y$ = # of children
$z$ = most amount spent

$$\left[(y-1) \times x\right] + 1 = z$$

In order to find what the most amount needed to be spent to satisfy given conditions, we must look at the worst case scenario. Because the objective is to find $y$ of a given color, the worst possibility is to get one less than $y$ of every color. Then causing an additional gumball to be purchased in order to obtain a $y^{th}$ matching color.
Argument 4

Ms. Hernandez came across a gumball machine one day when she was out with her twins. Of course, the twins each wanted a gumball. What’s more, they insisted on being given gumballs of the same color. The gumballs were a penny each, and there would be no way to tell which color would come out next. Ms. Hernandez decides that she will keep putting in pennies until she gets two gumballs that are the same color. She can see that there are only red and white gumballs in the machine. (The following questions are to help guide you towards answering question 1. You should complete all of the bulleted questions, but you only need to email complete responses for Questions 1-4 below. Based on your solution to Question 1 below, I may ask you to turn in your bulleted responses in class.)

- Why is three cents the most Ms. Hernandez will have to spend to satisfy her twins?
- The next day, Ms. Hernandez passes a gumball machine with red, white, and blue gumballs. How could Ms. Hernandez satisfy her twins with their need for the same color this time? That is, what is the most Ms. Hernandez might have to spend that day?
- Here comes Mr. Hodges with his triplets past the same gumball machine with red, white, and blue gumballs. Of course, all three of his children want to have the same color gumball. What is the most he might have to spend?

**Question 1.** Generalize this problem as much as you can. Vary the number of colors. What about different size families? Prove a generalization that always works for any number of children and any number of gumball colors. I encourage you to think about using alternative representations within your proof in order to create your most convincing/compelling argument.

\[
x - \text{represents # of colors} \\
y - \text{represents # of children} \\
z - \# of tries \\
x \cdot y = x + 1
\]

<table>
<thead>
<tr>
<th>PPI</th>
<th>colors</th>
<th>tries</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

Ex: If 4 kids x 3 colors (R, W, B)

\[
12 - 3 + 1 = 10
\]

- RWB
- RWB
- RWB
- RWB

equals 10

- RWB
- RWB
- RWB
- RWB

equals 10

add 1 color

4 colors of red
Argument 5

Ms. Hernandez came across a gumball machine one day when she was out with her twins. Of course, the twins each wanted a gumball. What’s more, they insisted on being given gumballs of the same color. The gumballs were a penny each, and there would be no way to tell which color would come out next. Ms. Hernandez decides that she will keep putting in pennies until she gets two gumballs that are the same color. She can see that there are only red and white gumballs in the machine. (The following questions are to help guide you towards answering Question 1. You should complete all of the bulleted questions, but you only need to enroll complete responses for Questions 1.4 below. Based on your discussion to Question 3 below, I may ask you to turn in your bulleted responses in class.)

- Why is three cents the most Ms. Hernandez will have to spend to satisfy her twins?
- The next day, Ms. Hernandez passes a gumball machine with red, white, and blue gumballs. How could Ms. Hernandez satisfy her twins for the same color this time? That is, what is the most Ms. Hernandez might have to spend that day?
- Here comes Mr. Hodges with his triplets past the same gumball machine with red, white, and blue gumballs. Of course, all three of his children want to have the same color gumball. What is the most he might have to spend?

Question 1. Generalize this problem as much as you can. Vary the number of colors. What about different size families? Prove a generalization that always works for any number of children and any number of gumball colors. I encourage you to think about using alternative representations within your proof in order to create your most convincing/compelling argument.

This is for positive integers only.

\[ C \times \text{# of colors to choose from} \]

\[ X \times \text{specific colors} \]

For any situation where you need \( X \) number of a single color of gumball with \( C \) colors available, there is an equation that tells you the maximum number of gumballs you’ll have to buy to get \( X \) of the random single color.

\[ \left( \frac{X-1}{C} \right) + 1 = \text{max number of gumballs needed} \]

The idea was to find an equation to show the max number of gumballs that wouldn’t satisfy the conditions and then add one. Add one because then you’d only need one more of any of the colors to have enough.

<table>
<thead>
<tr>
<th></th>
<th>A1 - Phillip</th>
<th>A2 - Keyshawn</th>
<th>A3 - Wendy</th>
<th>A4 - Monique</th>
<th>A5 - Penelope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrew</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>(no response)</td>
</tr>
<tr>
<td>Brenda</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Calvin</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Danielle</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Devin</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>(no response)</td>
</tr>
<tr>
<td>Elizabeth</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Gabriel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hilary</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>(no response)</td>
</tr>
<tr>
<td>Jeremy</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Joslyn</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Keyshawn</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Monique</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Martin</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Natasha</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(no response)</td>
</tr>
<tr>
<td>Penelope</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Peter</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Phillip</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Rachel</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rhonda</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Steven</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Stacey</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Tyrone</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Thomas</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Wendy</td>
<td>(no response)</td>
<td>(no response)</td>
<td>(no response)</td>
<td>(no response)</td>
<td>(no response)</td>
</tr>
<tr>
<td>Average</td>
<td>0.087</td>
<td>0.652</td>
<td>0.696</td>
<td>0.261</td>
<td>0.526</td>
</tr>
</tbody>
</table>

(Return to page 37)
Appendix C: Author Yee

Argument 1

First assign the variables.

Let $g$ be the number of colors of the gumballs.
Let $k$ be the number of children.
Let $y$ be number of pennies.

A general formula for this problem would be

$$y=g(k-1)+1.$$ 

This formula gives us the amount of most pennies needed to buy enough gumballs such that each child has the same color gumball.

To prove that this formula holds we can use an induction proof. This will be a two case proof. The first case is where the gumballs are fixed at a certain amount and the second case is where the children are fixed at a certain amount.

Case 1: Let $p(k)$ represent the equation $y=g(k-1)+1$ and let $g$ be fixed. Then we need to show that $p(k)$ is true for all $k$. For the basis step we show $p(1)$ is true, so $g(1-1)+1=1$. This is true, since there is only 1 child the matching criteria of the gumball colors is nullified so it will only take 1 penny to get a that child a gumball. Now we need to assume that $p(k)$ is true for all $k$ and prove $p(k+1)$ is true. If we apply $k+1$ we obtain $g(k-1)+1+g(k+1-1)+1=g(2k-1)+1 + g(2k)+2$, which is what we wanted and will be true. So by induction this equation is true for all sizes of families when the gumball colors are fixed.

Case 2 follows almost exactly the same except fixing the children and letting the amount of gumball colors vary.
Argument 2

<table>
<thead>
<tr>
<th># of Children</th>
<th># of Colors</th>
<th>Amount Spent (¢)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>19</td>
</tr>
</tbody>
</table>

From the table above we can see a pattern going on. When we have 2 children the amount spent goes up by one penny. When we have 3 children and increase the amount of colors the amount of pennies spent goes up by two. Example: When we have 4 children and 6 colors the first 6 pennies will give me one of each different color. The next 6 pennies will give me one of each different color again. This process is the same for the third time 6 more pennies will be spent. Until the fourth time we know we will get any color matching 4 color gumballs.

Amount Spent = A
Children = K
Colors = k
A = (K-1)(C)+1

When we have 2 children we see that the amount spent is simply C+1. But then when we have 3 children the amount spent is by two pennies. This is where the K-1 comes in place.
Argument 3

Let \( K \) be the number of kids in a family and let \( C \) be the color of gumballs in the machine. We need to find an equation that satisfies any number of kids to any number of gumball colors in the machine.

So if we have \( K \) kids minus 1 times the number of \( C \) colors plus 1 gives us the desired number of \( P \) pennies.

\[
\begin{align*}
  k & = \text{kids} \\
  c & = \text{color} \\
  p & = \text{pennies} \\
  \forall k, c, p \in \mathbb{N} \quad (k - 1)(c) + 1 = p
\end{align*}
\]

\( \text{Pf:} \)

Let \( k, c \) and \( p \) be arbitrary natural numbers. We need to show that we will always get an even or odd number for the value of \( p \). If we were to choose \( k \) to be either even or odd and \( c \) to be even or odd \( p \) will still be an even or odd number. We have four cases to check.

Case 1: \( k \) is even and \( c \) is even then \( p \) is odd.
Let \( k = 2n \) for all \( n \) in \( \mathbb{N} \) and \( c = 2m \) for all \( n \) in \( \mathbb{N} \), then

\[
\begin{align*}
  (2n - 1)(2m) + 1 &= p \\
  4nm - 2m + 1 &= p \\
  2(2nm - m) + 1 &= p \\
  2r + 1 &= p \quad \text{where} \quad r = 2nm - m \; \forall r \in \mathbb{N}
\end{align*}
\]

Therefore \( p \) is odd.

Case 2: \( k \) is odd and \( c \) is even then \( p \) is odd.
Let \( k = 2n + 1 \) for all \( n \) in \( \mathbb{N} \) and \( c = 2m \) for all \( m \) in \( \mathbb{N} \), then

\[
\begin{align*}
  ((2n + 1) - 1)(2m) + 1 &= p \\
  (2n)(2m) + 1 &= p \\
  2(2nm) + 1 &= p \\
  2r + 1 &= p \quad \text{where} \quad r = 2nm \; \forall r \in \mathbb{N}
\end{align*}
\]

Therefore \( p \) is odd.

Case 3: \( k \) is even and \( c \) is odd then \( p \) is even.
Let \( k = 2n \) for all \( n \) in \( \mathbb{N} \) and \( c = 2m + 1 \) for all \( m \) in \( \mathbb{N} \), then

\[
\begin{align*}
  (2n - 1)(2m + 1) + 1 &= p \\
  (4mn + 2n - 2m - 1) + 1 &= p \\
  4mn + 2n - 2m &= p \\
  2(2mn + n - m) &= p \\
  2r &= p \quad \text{where} \quad r = 2nm + n - m \; \forall r \in \mathbb{N}
\end{align*}
\]

Therefore \( p \) is even.

Case 4: \( k \) is odd and \( c \) is odd then \( p \) is odd.
Let \( k = 2n \) for all \( n \) in \( \mathbb{N} \) and \( c = 2m \) for all \( n \) in \( \mathbb{N} \), then

\[
\begin{align*}
  ((2n + 1) - 1)(2m + 1) + 1 &= p \\
  (2n)(2m + 1) + 1 &= p \\
  4mn + 2n + 1 &= p \\
  2(2mn + n) + 1 &= p \\
  2r + 1 &= p \quad \text{where} \quad r = 2nm + n \; \forall r \in \mathbb{N}
\end{align*}
\]

Therefore \( p \) is odd.

Therefore we showed that \( p \) will either be even or odd in the natural numbers, and \( p \) will give us the amount of coins need to get the same color.
Argument 4

Let $c$ be the number of colors of gumballs in the gumball machine. Let $k$ be the number of kids that want the same color gumball in a family. Then $g(c,k) = [(c)(k-1)]+1$ represents the maximum number of pennies a parent would have to spend to satisfy their kids.

Proof:
Let $c$ be the number of colors of gumballs in the gumball machine. Let $k$ be the number of kids that want the same color gumball in a family. We need to show that $g(c,k) = [(c)(k-1)]+1$ represents the maximum number of pennies a parent would have to spend to satisfy their kids. If we multiply $c$ by $(k-1)$ then either

1. Each color came out $(k-1)$ times and $(k-1)$ kids have all $c$ colors.
2. Or at least one color came out at least $k$ times, because another color came out less than $(k-1)$ times. If 2 happens then all of the kids have the same color gumball and they are happy. If 1 happens then one more kid needs a matching gumball. In this case the next gumball will satisfy the $k^{th}$ kid because the first $(k-1)$ kids already have all $c$ colors. Thus the most a parent would have to spend is $g(c,k) = [(c)(k-1)]+1$ colors.
Argument 5

If you want to have 2 gumballs that match, then the worst-case scenario to obtain 2 matching gumballs given n different colors possible is n+1 gumballs. To show this, it is clear that with up to n gumballs, it is possible for all n to be different colors, given n different colors. But by the pigeonhole principle, once you get the n+1th gumball, it has to match one of the other colors, and you have your 2 matching gumballs.

If you need m matching gumballs, it is a similar method. If you want to match m gumballs, given n colors, then clearly you can get a maximum of m-1 gumballs for each of the n colors, giving n(m-1) total gumballs. When you then obtain one more gumball, again, by the pigeonhole principle, you will obtain m of a certain color. Thus, for m matching gumballs given n total colors, the maximum number of gumballs needed is n(m-1)+1.

<table>
<thead>
<tr>
<th></th>
<th>A1 - Sonia</th>
<th>A2 - Vinny</th>
<th>A3 - Andy</th>
<th>A4 - Rose</th>
<th>A5 - Luke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andy</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Abby</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Connie</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Guadalupe</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>John</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Jacqui</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Luke</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Michael</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Antonella</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rose</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ryan</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sonia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stacey</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tanya</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Elizabeth</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Vinny</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Average</td>
<td>0.5</td>
<td>0.25</td>
<td>0.5625</td>
<td>0.75</td>
<td>0.6825</td>
</tr>
</tbody>
</table>

(Return to page 37)
Appendix D: Author Ko

Argument 1

Children = c
Gumballs = g
Pennies = p
P = g (c - 1) + 1

In order to make sure you get enough gumballs for every child to have one of the same color you must buy one of all the colors for every child but one. Then no matter what color you get next every child will have one of the same color. Therefore multiply the number of gumball colors by one less than the number of children, after that add one.

2c and 2g = 3p
2c and 3g = 4p
2c and 4g = 5p
3c and 2g = 5p
3c and 3g = 7p
7c and 2g = 13p
Argument 2

<table>
<thead>
<tr>
<th>(c) (number of children)</th>
<th>(g) (number of colors of gumballs)</th>
<th>(f(c, g)) (maximum cost)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>10</td>
</tr>
</tbody>
</table>

After reading and thinking through the given scenario, I believe it is best proved by generalizing the examples that were given. We know that the maximum cost of the gumballs will depend on both the number of children AND the number of colors in the machine. The maximum cost will measure the worst case possible: when the family receives all colors of gumballs before receiving repeats of colors. Therefore, for this proof we will refer to a group of gumballs that contains one and only one gumball from each color as a “set”. Thus in the worst case possible, a family will accumulate one set of gumballs before receiving repeats of colors. Understanding this, if a family accumulates \((c-1)\) sets, the next gumball purchased from the machine will allow for each child to have the same color because each previous set will include the color that comes out as well. Thinking about this in terms of a function, \(f(c, g) = [g^*(c-1)] + 1\). In this function, \([g^*(c-1)]\) represents the total number of gumballs that will be in all the \((c-1)\) complete sets when added together. Therefore, the \(+1\) represents the next gumball that will allow all of the children to have the same color. Because each gumball is one penny, the total number of gumballs purchased will equal the amount of money that is spent.

Visual representation: \(f(c, g) = [g^*(c-1)] + 1\)

\[c = \text{number of children}\]  \[g = \text{number of colors}\]
Argument 3

In order to generalize this problem we need to think about it flexibly with varying colors and children.

Visualize:

Ex: 3 children 5 colors:

colors: 1 2 3 4 5

In order to achieve 3 sets of the same color, 2 rows of colors should be filled up, then 1 for the final color to make a set of three for one color.

So:

on this penny, we now have 3 of the same color. Since the number of rows depends on the number of children, then we can say that there are n-1 since we do not need the last row filled up completely (that would give us a 3 set of the same color for each color). We take away one for the rows and we need to multiply this by the # of colors of gumballs to get the number of gumballs that we will recieve and then add one more to represent the final gumball in the last row. So the formula should look something like this:

\[(\text{color of gumballs} \times \text{rows}) + 1 = \text{#} \text{ of gumballs received}\]

\[c (n-1) + 1 = p \text{ where } c = \text{colors} \text{ and } p = \text{pennies.} \]

In order to prove this we need to look at the information:

For 5 different colors and n # of children we have:

\[
\begin{array}{c|c|c}
\text{colors} & \# \text{ of sets} & \text{# of gumballs received} \\
\hline
1 & 5 & \\
2 & 10 & \\
3 & 15 & \\
4 & 20 & \\
5 & 25 & \\
\end{array}
\]

So we can see that if we need 3 sets of the same gumballs colors we would need 15 gumballs.

The relation ship would be $3 \times (5-1) = 3\times4 = 12$ where n is number of children so

$5k = \# \text{ of gumballs received}$

So:

Where $c = \# \text{ of colors then}$:

\[
\begin{array}{c|c|c}
\text{colors} & \# \text{ of sets} & \text{# of gumballs received} \\
\hline
1 & 5 & 15 \\
2 & 10 & 30 \\
3 & 15 & 45 \\
4 & 20 & 60 \\
5 & 25 & 75 \\
\end{array}
\]

Since we only need one extra gumball to achieve our goal we add one more gumball:

\[ck + 1 = \text{final number of gumballs}\]
Argument 4

Proof: \( m = kc - (c-1) \)

Let \( m \) = amount of change it cost
\( c \) = colors of gumballs
\( k \) = amount of children

By spending 1 cent a R or W ball will be chosen. The next 1 cent will represent the ball in the opposite color. The third 1 cent will be either R or W, but regardless it will be a match to one of the previously chosen colors. Thus, the maximum amount of money being spent by Ms. Hernandez for her twins will be 3 cents.

By spending 3 cents she will receive three gumballs. They must be either red or white, and therefore there must be at least two of the same color. The possible outcomes of receiving three gumballs is: 3 red, 3 white, 2 red 1 white, or 1 red 2 white. Either way there will be two of the same color for her twins.

<table>
<thead>
<tr>
<th>( X )</th>
<th>( Y )</th>
<th>( R )</th>
<th>( W )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1 cent</td>
<td>1 cent</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1 cent</td>
<td></td>
</tr>
</tbody>
</table>

This same idea will work for however many colors and amount of children. For three kids and three different colors, spending 6 cents will give two of every color and the next 1 cent will give a match to any of the three colors. Thus, a correlation can be made to say the maximum amount of money that will be spent will be the amount of kids times the amount of colors. Then subtract the amount of colors-1 from the product. This color minus 1 results from not having to have all the possible options. 3 kids and 3 colors will be 9 for the amount of money being spent, but since just spending 7 will give three of one of the colors, 2 can be subtracted from the 9. This leads to \( c-1 \) being a part of \( m = kc - (c-1) \).
Argument 5

Conjecture: For any number of kids, k and any number of gum ball colors, c, where each kid must have the same gum ball color the total worst case scenario cost 1c per gum ball is c(k-1) + 1

Assume the same color is drawn for each kid except one kid because if every kid got the same color it would not be the worst possible case. Then a new streak of color (2nd color) is drawn for every kid but again it stops one short to match every kid. Then a third color begins to be drawn from the machine and again the number of gum balls of this third color is one less than is needed.

Here let me show you a picture of what I mean.

<table>
<thead>
<tr>
<th></th>
<th>A1 - Brandon</th>
<th>A2 - Dennis</th>
<th>A3 - Gina</th>
<th>A4 - Henry</th>
<th>A5 - Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brandon</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Carla</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dennis</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Francine</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Greg</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Gina</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Henry</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Joel</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Jessica</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Katherine</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Lasondra</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Naomi</td>
<td>(no response)</td>
<td>(no response)</td>
<td>(no response)</td>
<td>(no response)</td>
<td>(no response)</td>
</tr>
<tr>
<td>Average</td>
<td>0.307692308</td>
<td>0.769230769</td>
<td>0.846153846</td>
<td>0.153846154</td>
<td>1</td>
</tr>
</tbody>
</table>

(Return to page 38)
Enhancing Teachers’ Assessment of Mathematical Processes Through Test Analysis in University Courses*

Patricia D. Hunsader  
*University of South Florida, Sarasota-Manatee*

Barbara Zorin  
*St. Petersburg College*

Denisse R. Thompson  
*University of South Florida*

Assessment is a critical component of the teaching and learning cycle. Yet, research suggests that teachers have often had insufficient preparation relative to the development and use of assessment. In this article, we share experiences and assignments we use with both preservice and in-service teachers within undergraduate and graduate university courses to enhance their focus on mathematics assessment, particularly assessment of processes and practices in classroom tests. We also share the results of teachers’ analyses of classroom tests, their reactions to their analysis, and their reflections on the potential impact of the experiences on their future practice.

**Key words:** Mathematics assessment; Mathematical processes; Test analysis; Mathematics methods courses

Although classroom assessment is key to maximizing student learning (Lukin, Bandalos, Eckhout, & Mickelson, 2004), research indicates that teachers are often not adequately prepared to make decisions about classroom assessment practices (Fan, Wang, & Wang, 2011; Koh, 2011; McMillan, Myran, & Workman, 2002; Quilter & Gallini, 2000; Stiggins, 2002), and they recognize these deficits in their preparation (Chelsey & Jordan, 2012; Mertler, 2009; Zhang & Burry-Stock, 2003; Zientek, 2007). Assessment courses are not required in all teacher education programs, and such courses are often not content specific (DeLuca & Klinger, 2010), providing little to no instruction in mathematics assessment.

Numerous educators recommend that mathematics assessment address more than just knowledge of content. It should also provide insights into students’ higher-order thinking, reasoning, problem solving, communication, and conceptual understanding (Borko, Mayfield, Marion, Flexer, & Cumbo, 1997; Boston & Smith, 2009; NCTM, 1989, 1995, 2000). Because how students learn (i.e., the mathematical processes/practices through which they learn) is as important as what they learn (i.e., the specific mathematical content) (Kilpatrick, Swafford, & Findell, 2001; Ma, 1999; National Governors Association Center for Best Practices (NGA Center) and Council of Chief State School Officers (CCSSO), 2010; NCTM, 1989, 2000, 2006), assessments should focus not just on content mastery and procedural skills, but also on mathematical processes and practices that provide opportunities for students to demonstrate their mathematical thinking. As students provide insight into their thinking about important concepts and processes, teachers are positioned to modify and adjust classroom instruction to enhance learning.

Classroom assessment encompasses a range of forms, including informal observation, students’ individual and group work, performance assessments, portfolios, journal writing, and tests. The National Research Council, in its report *Everybody Counts* (1989), highlights one form of classroom assessment that deserves attention: “We must ensure that tests measure what is of value, not just what is easy to test” (p. 70). Tests also need to “be sensitive enough to help teachers identify individual areas of difficulty in order to improve instruction” (NCTM, 1989, p. 1054).

---

* This manuscript was edited by Margaret Smith, editor of the *MTE* at the time the manuscript was initially submitted.
Throughout, we refer to our university students as teachers, regardless of whether they are preservice or in-service. We use the term teachers to refer to their K–12 students.

Rationale for Enhancing the Focus on Test Analysis Relative to Mathematical Processes

Smith and Stein (1998) note that to evaluate a task effectively, it is necessary to look beyond the task’s surface features to analyze the kind of thinking required. Similarly, we believe it is important to look beyond the content being assessed by a test item to determine the extent to which it provides students with opportunities to engage with mathematical processes/practices teachers want to support. Test items may assess important content without focusing on mathematical processes but can often be easily modified to focus on processes while still assessing the intended content (e.g., Thompson, Beckmann, & Senk, 1997).

In our earlier research related to assessment, we built on the work of Senk, Beckmann, and Thompson (1997) and other researchers (e.g., Madaus, West, Harmon, Lomax, & Viator, 1992; Taylor, 1992) to develop a framework to analyze tests accompanying published curricula for the extent to which the test items provide potential opportunities for students to engage in important mathematical processes/practices. The language of our Mathematical Processes Assessment Coding (MPAC) framework is closely aligned to the language of the NCTM Process Standards (NCTM, 2000), providing content validity. The reliability of our framework was established for research purposes with coder agreement ranging from 90% to 98% across the framework criteria. (See Hunsader, Thompson, & Zorin, 2013, for information about the development of the framework.)

The results of our analysis of hundreds of chapter/unit tests accompanying published curricula from a range of publishers across elementary, middle, and high school grades indicate that such tests infrequently and inconsistently provide opportunities for students to engage in important mathematical processes and practices, such as reasoning, communication, connections, and representation (Hunsader et al., 2014; Hunsader, Thompson, & Zorin, 2013). These findings raised concerns that teachers primarily using the tests accompanying their curricula might not be preparing students for some of the new high-stakes assessments, such as our own Florida State Assessment, or the assessments associated with the Partnership for Assessment of Readiness for College and Careers (PARCC, 2013) or Smarter Balanced (2013). Furthermore, consistent with findings of researchers around the world (e.g., Blok, Otter, & Roeleveld, 2002; Delandshere & Jones, 1999; Garet & Mills, 1995; Madaus et al., 1992), we had documentation showing that many of the teachers in our state use the tests accompanying their curricula and are sometimes required to do so. Hence, we felt a need to support the development of teachers’ assessment knowledge relative to mathematical processes/practices by helping them become critical consumers of the tests they use with their students and by raising their awareness of the potential lack of coherence between those tests and the high-stakes tests their students are required to take.

Assessment Experience in Our Courses

Context

The assessment experience described here has been refined over multiple semesters within undergraduate and graduate courses for preservice and in-service teachers at all levels, K–12.¹ The undergraduate preservice teachers were enrolled in the first of two elementary mathematics methods courses. The first course focuses on number

¹ Throughout, we refer to our university students as teachers, regardless of whether they are preservice or in-service. We use the term students to refer to their K–12 students.
and operations, including work with whole numbers, fractions, and decimals, while the second course focuses on geometry, measurement, and basic statistics. At this point in their program, most teachers have participated in their first field experience but have had few opportunities to teach a lesson and have no experience in creating or modifying a test. The graduate teachers were enrolled in a master’s-level methods course on the teaching of algebraic ideas in elementary or middle grades or in a trends course looking at current issues in mathematics education; all teachers in the methods course were in-service teachers, but those in the trends course were either preservice teachers in a Master of Arts in Teaching program or in-service teachers. Typical class sizes range from 20 to 30 for the undergraduate courses and from 15 to 20 for the graduate courses. The lead author was the instructor for the undergraduate courses, and the other authors were instructors for the graduate courses.

At our institution, both undergraduate and graduate programs require a measurement course, which focuses on general issues about test development, including a focus on assessing course objectives and designing tests. Therefore, in our work with teachers we have focused on investigating the extent to which the tests include opportunities for students to engage with mathematical processes and have not focused on other aspects of test development.

**Describing Our Work with Teachers**

**Overview of the assessment experience.** The assessment experience consists of several components completed over multiple class sessions. The general structure used in the undergraduate and graduate methods courses is outlined in Table 1, together with approximate time allocations for each component; the structure for the experience in the trends course is described later. A list of potential readings is found in Appendix A: a detailed discussion of the components follows the table.

**Providing a background on standards and the language of assessment (Components I, II, and III).** We have found it helpful for teachers to have some background knowledge of the Standards movement, and such background is discussed early in the semester and then referenced.

---

**Table 1**

*Components of the Assessment Experience in Methods Courses*

<table>
<thead>
<tr>
<th>Time frame</th>
<th>Component</th>
<th>Undergraduate Methods (preservice teachers)</th>
<th>Graduate Methods (in-service teachers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5–2 hours</td>
<td>I. History of math standards</td>
<td>Overview of Standards from NCTM (1989) through Common Core State Standards (CCSSM, 2010) with emphasis on the NCTM Process Standards and the CCS Standards for Mathematical Practice</td>
<td>Overview of Standards as needed, with focus on process standards and standards for mathematical practice</td>
</tr>
<tr>
<td>0.5–1 hour</td>
<td>II. Readings on assessment</td>
<td>Assessment chapter from methods text (outside of class)</td>
<td>Research or practitioner journal articles about assessment (jigsaw in class)</td>
</tr>
<tr>
<td>0.25–0.5 hour</td>
<td>III. Introduction to language of assessment</td>
<td>Review assessment of and for learning, formative and summative roles of assessment, various forms of assessment (e.g., chapter tests, projects, journals, observations)</td>
<td></td>
</tr>
<tr>
<td>0.75–1 hour</td>
<td>IV. Introduction to our framework and coding process</td>
<td>Review framework codes and coding of sample items, engage teachers in whole-group coding of items, discuss possible modifications of items</td>
<td></td>
</tr>
<tr>
<td>0.5–1 hour</td>
<td>V. Small-group analysis</td>
<td>Small groups code and modify sample items and present their items, codes, and suggested modifications to the class for discussion and consensus</td>
<td>Small groups code sample items or 1–2 complete tests, develop possible modifications; whole-class discussion of small-group work</td>
</tr>
<tr>
<td>Varies</td>
<td>VI. Independent assignment</td>
<td>Analyze a collection of sample items and provide modifications as needed</td>
<td>Analyze complete tests and modify items as needed</td>
</tr>
</tbody>
</table>

*Note.* In the undergraduate methods course, Component I is discussed during the first week of class and Components III, IV, and V are discussed the 3rd week. For the graduate methods course, Component I is discussed during the first week as needed, with Components II, III, IV, and V covered during the 6th week of the 15-week semester.
during the assessment experience. To introduce teachers to the language of assessment, we provide a reading assignment that varies based on teachers’ backgrounds. Preservice teachers read the portion of their methods text that introduces assessment. For in-service teachers, we add readings about classroom assessment appropriate to the teachers’ grade level (see Appendix A).

To help set the stage for the analysis of tests, we discuss the Common Core State Standards for Mathematics (CCSSM; 2010), specifically, the Standards for Mathematical Practice, and the standards that served as the foundation of our framework, the NCTM Process Standards (NCTM, 2000) of problem solving, reasoning and proof, communication, connections, and representation. In particular, we discuss the power of mathematical processes and practices to elicit student thinking and reveal students’ level of understanding of mathematical concepts, with the intention that such thinking provides teachers with insight into student learning that can help them modify instruction.

As part of our introduction, we also discuss that assessment is both of and for student learning, and that assessments can serve both formative and summative roles depending on how the information from the assessment is used (Joyner & Muri, 2011). We discuss various forms of assessment that are typically used in classrooms, and how each form serves a different purpose. For instance, formative assessment often includes journal writing, which provides opportunities for students to explain their thinking or describe how they are feeling about their learning; observations (including facial expressions), which give in-the-moment insight into students’ understanding; and bellwork, which enables misconceptions to be caught early or prerequisite knowledge to be foregrounded. Summative assessments can include tests as well as projects that provide opportunities for students to engage in a problem over an extended period of time. Although mathematics teachers are unlikely to eliminate tests in the foreseeable future, nor do we suggest such a practice, we want teachers to realize there are many opportunities for assessment beyond paper-and-pencil tests, even though tests are the main focus of our assessment experience.

Providing a tool to analyze tests objectively (Component IV). Because we had previously developed the MPAC framework for analyzing mathematical processes on tests accompanying published curricula and used it to analyze hundreds of tests, we believed it would be a useful tool for teachers to analyze tests in an objective manner. Figure 1 illustrates our framework and the criteria used to analyze assessments for Reasoning and Proof, Communication, Connections, Role of Graphics, and Translation of Representational Forms. Figure 2 includes sample items we created for the purposes of this article to illustrate distinctions among the framework codes; the items span elementary to high school content and a range of topics. The codes for the sample items are provided in Table 2. For additional sample items and explanations of coding, see Hunsader et al. (2013, 2014).

Our criteria of Reasoning and Proof and Communication are related; it is possible for an item to elicit students’ mathematical communication without engaging them in reasoning and proof, but all instances of reasoning and proof involve communication. Item 1 only directs students to record a numeric answer (Communication code N). Two different modifications of item 1 would have resulted in an affirmative code for Communication: if the item did not include a graphic but asked students to draw one (graphic communication), or if the item included the graphic but no equations and asked students to write the related multiplication and division equations illustrated by the graphic (symbolic communication). Items 3 and 4 are coded as involving Reasoning and Proof because both require students to provide a justification for their answer. Although item 2 includes a great deal of Communication (code Y) in the writing of equations, students are not asked to justify their responses.

We distinguish between these two criteria as follows: Communication involves students’ sharing what they are thinking; Reasoning and Proof goes beyond that to elicit students’ rationale for why they gave that response. (We originally developed the MPAC to analyze elementary tests, for which the term justification seemed softer than proof; as we expanded our work to middle and high school grades, we kept the term justification, recognizing that what counts as justification becomes more rigorous across the grades. We believe this description is in line with an essential principle related to all proof: “To specify clearly the assumptions made and to provide an appropriate argument supported by valid reasoning so as to draw necessary conclusions” [Hanna et al., 2009, p. 19].)

All four items are coded as involving Connections. Mathematical connections can take two major forms: real-world connections (items 2 and 4), and interconnections of mathematical concepts (items 1 and 3). Item 1 relates multiplication/division and item 3 relates ideas of parallel lines, corresponding angles, and similar triangles. In our work with test-item analysis, we have rarely found items set in a real-world context that also interconnect mathematical concepts. In those cases, we coded the item as R but acknowledge that other users may prefer to indicate both the R and I codes.

Students use Representation in various ways to express their mathematical thinking (Diezmann & McCosker,
Role of Graphics captures how students interact with the graphics to provide a successful response (Berends & van Lieshout, 2009). When students must interpret (item 4) or create (item 3) a graphic, it is clear the item involves graphics. In contrast, some items include a graphic, but there is enough information provided in the stem of the item that a student could ignore the graphic and still respond correctly. In item 1, the counters are arranged in such a way that the mathematical relationship between 2 rows with 6 in each row and 12 counters broken into two sets is apparent. In this case, although the graphic is not needed to answer the question, it clearly represents the mathematics in the item (code R). If item 1 omitted the equations and asked students to create equations based on the graphic, it would require interpretation of the graphic and receive a code of I. If item 2 had included a picture of some birds and lizards, we would have coded it S.

In Translation of Representational Forms, we look for items that involve multiple representations by asking students to record a translation from one form of the item (symbols, words, or graphics) to a different one, a
1. Raul placed his counters in an array.

What number completes these related facts?

\[ 2 \times \_ = 12 \]

\[ 12 \div 2 = \_ \]

2. Walton has 14 pet birds and lizards. Together, his pets have a total of 44 legs. How many of Walton’s pets are birds? Solve using a system of equations. Record both equations, and the answer.

First equation: 

Second equation: 

Number of birds: 

3. Prove that the segment joining the midpoints of sides \( AB \) and \( AC \) in triangle \( ABC \) is parallel to side \( BC \) and half its length. Draw a diagram to support your work.

4. Jared, Rayvon, and Chou are the top athletes on their track and field team. The table below shows how they placed in 5 events for the last track meet.

<table>
<thead>
<tr>
<th>Event</th>
<th>100 m</th>
<th>200 m</th>
<th>400 m</th>
<th>800 m</th>
<th>1500 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jared</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Rayvon</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Chou</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

For an upcoming meet involving sprints (100 m, 200 m, and 400 m) and medium distances (800 m and 1500 m), the school can only send two athletes. Based on these results, which two should the school send? Justify your decision.

Table 2

<table>
<thead>
<tr>
<th>Item</th>
<th>Reasoning &amp; Proof</th>
<th>Communication</th>
<th>Connections</th>
<th>Representation: Graphics</th>
<th>Representation: Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td>N</td>
<td>I</td>
<td>R</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>N</td>
<td>Y</td>
<td>R</td>
<td>N</td>
<td>SW</td>
</tr>
<tr>
<td>3</td>
<td>Y</td>
<td>Y</td>
<td>I</td>
<td>M</td>
<td>WG</td>
</tr>
<tr>
<td>4</td>
<td>Y</td>
<td>Y</td>
<td>R</td>
<td>I</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 2. Sample items to illustrate framework codes.

critical skill in developing deep understanding of concepts (Bossé, Adu-Gyamfi, & Cheetham, 2011; Gagatsis & Shiakalli, 2004). In making this determination, it helps teachers to consider what would be involved in translating an item from English to Spanish. That translation would need to capture all the critical elements of the problem in the second language, far beyond merely providing an answer in the second language. Item 1 in its original form does not require a translation. However, if it omitted the equations and asked students to write equations to represent the graphic, or omitted the graphic and asked students to create a graphic to represent the equations, we would have recorded a code of \( G5 \), indicating a translation between graphic and symbolic representations. Item 2 is coded as \( SW \); to complete the task, students must translate the word problem into symbolic equations. Item 3 involves translating the words of the problem into a graphic (code \( WG \)).

Because our assessment work began as a research project analyzing tests provided by textbook publishers without considering the written instruction in the text or the
classroom experiences of students, we did not include problem solving in the framework. A determination of whether an item might engage students in problem solving cannot be made without understanding students’ background knowledge or the instruction they have experienced; however, it is clear that many procedural items (e.g., $2.38 \times 4.76 = ?$ for middle grades students) would likely provide few opportunities for problem solving regardless of classroom instruction. Although we have continued to exclude problem solving from our framework and our initial analysis of tests, we revisit this issue in the discussion.

**Practice coding and modifying items (Component V).**
After introducing teachers to the MPAC framework and its relationship to the NCTM process standards, teachers work in a whole-class activity to create a shared understanding of the framework criteria and codes. Teachers code a selection of items or a whole test, discuss challenges and disagreements in coding, and discuss possible ways to modify items so that potential opportunities to engage with the mathematical processes/practices become evident. The author team creates the items, which are representative of items on published tests we have analyzed and involve a variety of framework codes. We discuss the risk of analyzing individual items in isolation and the goal of integrating a range of processes/practices across an entire test rather than attempting to embed all processes/practices within each item. After whole-class coding, small groups work to analyze and modify more items or tests. They then share their work with the whole class for additional discussion.

**Independent assessment assignment (Component VI).**
As a follow-up to the class experience, teachers complete an independent assignment related to analyzing tests or test items. Undergraduate preservice teachers are given a set of 5–8 items created by the author team to model a range of framework codes across grade levels with content that varies based on the content of the course. For each item, teachers record a code for each framework criterion and a rationale for their selection and then suggest a modification to the item as needed. At the end of the assignment, they reflect on their use of the framework and how this work might impact how they formulate chapter tests in their future classrooms.

Graduate in-service teachers bring a different perspective to the experience because they are typically working with students whose abilities and challenges they understand. Such teachers have taught content units and then assessed their students’ mastery of that content. Regardless of whether they have used their own tests or those provided either by a publisher or their district, these teachers have graded such tests and observed what they reveal about their students’ understanding. To increase the relevance of the assignment for these teachers, we have them analyze two to five whole tests of their choosing, whether self-authored or provided by their textbook publisher or school district. We challenge teachers to reflect on their personal beliefs about the appropriate level of mathematical processes/practices across an entire test. We do not suggest a target level; that decision must rest with a classroom teacher who knows her/his students and what other assessment measures are used. However, we require teachers to modify at least two items strategically so that the processes/practices become more evident.

**Evaluation of the independent assignment.** The emphasis placed on the independent assignment for a course grade is typically about 10% in the undergraduate and graduate methods courses based on the amount of class time dedicated to assessment and the overall expectations for the assignment. Later in this article, we share an adaptation/extension of the assignment for a course in which more time is allocated to assessment.

Our grading criteria are not focused on whether teachers apply our framework correctly or incorrectly but rather on their rationale for their analysis and modification of items and their reflection on their own assessment practices. Such a grading decision is consistent with our intent to enhance teachers’ ability to be critical consumers of their tests, rather than suggest that there is a perfect test or a predetermined ratio of processes/practices that should be present in tests.

Figure 3 shows two items that preservice teachers analyzed, along with their codes and rationales. We evaluated the code and rationale holistically, and the examples are ones that would have received partial credit: either a code different from what we assigned is given but with a reasonable rationale, or a code consistent with ours is given but the rationale is lacking in some respect.

Teacher responses A and C include a code that is different from how we coded the item (V and R, respectively), but the rationales demonstrate good thinking about how students might engage with the items. For item 1, Teacher A understood that students were not expected to explain their thinking but missed that the item asks students to interpret a representation of vocabulary. For item 2, Teacher C makes an incorrect assumption that students are expected to complete the graphic but chooses the correct code based on that assumption.

Teachers B and D coded the item as we expected, but their explanations do not fully communicate a rationale for the code. Teacher B’s rationale incorrectly implies that the student has to write a response including vocabulary.
Item 1: Tell if the line is a line of symmetry.

Framework Criterion: Communication

Teacher A: Code N. No communication is represented in this question. Students do not have to show their thought processes, nor does it engage students in the use of mathematical language.

Teacher B: Code V. The test item directs students to communicate their mathematical thinking solely through vocabulary.

Item 2: Samantha has 40 dolls. She puts them in 5 equal groups. How many dolls are in each group?

<table>
<thead>
<tr>
<th>40 dolls in all</th>
</tr>
</thead>
</table>

Framework Criterion: Role of Graphics

Teacher C: Code M. A partial graphic is given and the item directs students to add to the existing graphic. Students fill in the blanks where there are question marks.

Teacher D: Code R. A graph is given, and it would be helpful for a student to use it. There is no direction on how to use it, or that they need to use it to show work.

Figure 3. Sample test items analyzed by preservice teachers with their codes and rationales.

Teacher D acknowledges that the graphic may support students but does not communicate awareness that the graphic completely illustrates the mathematics in the problem.

Teacher Reflections on the Assessment Experience and the Framework

We continue to be interested in teachers’ perspectives on the assessment experience embedded within our methods courses so we can improve the overall experience, maximize its potential to impact teachers and their students, and ensure that the time spent is perceived as worthwhile by the teachers. Thus, after implementing the assessment experience in several course sections, we began a formal research study to collect data about teachers’ perspectives on the use of the framework and their perceptions about how the analysis of tests/test items might influence their future classroom assessment practices. We report data only from those sections in which teachers were completing the assessment experience for the first time, provided informed consent for the use of their work, and for which we have all their work electronically.

Reflections from undergraduate preservice teachers.

Preservice teachers responded to the following questions:

- To what extent did the framework impact your ability to analyze whether a test item gave students an opportunity to engage in the mathematical processes you evaluated?
- How might these assessment experiences impact your formulation of chapter tests in mathematics when you are a teacher?

Preservice teachers’ responses were analyzed according to grounded theory and emergent design (Glaser & Strauss, 1999; Patton, 2002). Responses to the two questions from 23 preservice teachers in the elementary methods course were coded and discussed by the authors using a consensus model (Charalambous, Delaney, Hsu, & Mesa, 2010; Ding & Li, 2010; Roseman, Stern, & Koppal, 2010) to identify segments that seemed to indicate a coherent thought. We identified and coded 72 thought segments among the responses and then collapsed related codes into six themes, labeled as indicated.
below, along with the percentage of teachers who had at least one thought segment that fell into that theme.

A: Comments about framework criteria and what teachers need to use the framework effectively (83% of teachers)

B: What the framework affords the user (57% of teachers)

C: Quality of framework criteria (57% of teachers)

D: Value of analysis for modification of items and building of quality tests (13% of teachers)

E: Comments on the overall assignment (17% of teachers)

F: Other (9% of teachers)

Table 3 illustrates one teacher’s response to the two questions, the thought segments we coded, and the theme in which that code was eventually placed.

We discuss teachers’ reflections for themes A–D, leaving discussion of theme E to a later section of this article.

Theme A: Comments about framework criteria and what teachers need to use the framework effectively. Eighty-three percent of the teachers made comments about some aspect of the framework criteria and supports that would help in using it. Some reflections detailed specific

<table>
<thead>
<tr>
<th>Original Response</th>
<th>Segment of Thought</th>
<th>Code</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>I found the way that the MPAC framework was set up was helpful. It was organized</td>
<td>I found the way that the MPAC framework was set up was helpful. It was organized</td>
<td>Organized</td>
<td>B: Framework affords</td>
</tr>
<tr>
<td>well so that I could assess the problems easily. Under each term, like Reasoning</td>
<td>well so that I could assess the problems easily. Under each term, like Reasoning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and Proof, it said Justification, which gave me a quick clue to make the assessment</td>
<td>and Proof, it said Justification, which gave me a quick clue to make the assessment</td>
<td>Clear criteria</td>
<td>C: Quality of criteria</td>
</tr>
<tr>
<td>accessible. I have never done any sort of assessments on student problems… I really</td>
<td>accessible. I have never done any sort of assessments on student problems… I really</td>
<td></td>
<td></td>
</tr>
<tr>
<td>liked having to try to correct the problems (trying to make them better). I felt</td>
<td>liked having to try to correct the problems (trying to make them better). I felt</td>
<td>Value of modifications</td>
<td>D: Aids in modifications and building of quality tests</td>
</tr>
<tr>
<td>that it helped me make more sense of what was wrong with the problem in the first</td>
<td>that it helped me make more sense of what was wrong with the problem in the first</td>
<td></td>
<td></td>
</tr>
<tr>
<td>place. (T215 response to question 1)</td>
<td>place. (T215 response to question 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The only thing I found challenging about the MPAC-5 framework was that I was second-</td>
<td>The only thing I found challenging about the MPAC-5 framework was that I was second-</td>
<td>Struggle with criteria</td>
<td>A: Comments about criteria and what teachers need to use the framework effectively</td>
</tr>
<tr>
<td>guessing myself on certain codes. Some items seemed to be able to be one code or</td>
<td>guessing myself on certain codes. Some items seemed to be able to be one code or</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the other (it seemed to be both!). At first I had a hard time figuring out if a</td>
<td>the other (it seemed to be both!). At first I had a hard time figuring out if a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>problem had translation or asked the student to translate and then when I read</td>
<td>problem had translation or asked the student to translate and then when I read</td>
<td></td>
<td></td>
</tr>
<tr>
<td>more about translation I was more comfortable coding. (T215 response to question</td>
<td>more about translation I was more comfortable coding. (T215 response to question</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2)</td>
<td>2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
criteria teachers found challenging (e.g., distinguishing between Reasoning and Proof and Communication); others provided suggestions that might support future work with the framework or acknowledged that analyzing assessment items was an entirely new experience. Typical responses include:

“I found that it may be helpful to provide a sample question to each of the coding frameworks [criteria] to help . . . differentiate between them.” (T221)

“I also found it difficult to complete this assignment because I have never thought about nor had [to] practice assess [sic] test questions.” (T226)

Theme B: What the framework affords the user. Fifty-seven percent of the teachers made comments that we interpreted to indicate that they were thinking beyond the framework criteria to what the analysis reveals about the items and their ability to reveal student thinking, as suggested in the following representative comment:

“The [framework] made it very obvious that all of the questions required little to no thinking . . . If there was thinking required, students did not need to show it . . . It is easier to see a mistake when a student’s work is shown.” (T211)

Theme C: Quality of criteria. Fifty-seven percent of the teachers made comments that focused on how the structure and wording of the framework criteria helped in their analysis, while others noted a desire to internalize the framework criteria to guide their future assessment practices. The following comment suggests how one teacher considered turning the criteria into questions to ask herself when developing tests:

“The codes served as the questions that we (should) innately ask ourselves when sorting through these types of questions. They served as question prompts for those of us that are still digesting the information and trying to make sense of it.” (T205)

Theme D: Analysis aids modification of items and building of quality tests. Thirteen percent of teachers made comments that suggested a growing awareness of the complexity of crafting quality assessments, as reflected in the following comment:

“I never realized that so much went into writing a test question.” (T226)

Reflections from graduate in-service teachers. The in-service teachers in the graduate methods course reflected on the following three questions, with the first and third questions identical to those asked of the preservice teachers.

- To what extent did the framework impact your ability to analyze whether a test item gave students an opportunity to engage in the mathematical processes you evaluated?
- After using the framework to analyze tests, what did you notice and what conclusions did you draw about the tests you analyzed?
- How might these assessment experiences impact your formulation of chapter tests in the future?

Responses from the 19 graduate teachers were coded into 83 thought segments that were collapsed into the following themes using the same process we used for the preservice teachers’ responses.

G: Concern about test quality (100% of teachers)

H: Impact of assessment experiences on future assessment practices (84% of teachers)

I: Framework facilitates analysis of tests (47% of teachers)

J: Other (11% of teachers)

Again, we briefly discuss responses within these themes.

Theme G: Concern about test quality. Although these teachers have experience assessing their own students, for some this was their first experience analyzing the tests they used. All teachers made some comment that fell within this theme, with most indicating some level of disappointment with their tests after the analysis. The following comment suggests this teacher realized that her tests did not provide opportunities to elicit student thinking:

“I noticed that the items that I analyzed did not aptly allow the students to reason or justify their thinking. . . . I feel like that does the student a real disservice because I can’t truly interpret the data and drive my instruction if I don’t know the reasoning or provide my students with the opportunity to justify their thinking.” (T606)

Theme H: Impact of assessment experience on future assessment practices. The focus on test analysis was designed to heighten teachers’ awareness of the importance of including processes and practices within test items and to impact their future assessment practice. Reflections
from 84% of the teachers appear to indicate they recognize the potential negative impact on students’ learning when they limit assessment to the tests they currently use. In contrast to the work cited by Black and Wiliam (2010), these typical responses indicate an awareness of the need to alter future practices and a willingness to engage in the challenging work required to effect change:

“In the future, prior to administering a math test with students, an analysis of the test questions will be conducted using the . . . framework . . . to ensure that the mathematics concept and processes are being evaluated adequately.” (T617)

“The coding and reviewing of assessments has given me a lot to think about. Do I even bother looking at publisher based assessment or do I develop my own from scratch every time? The answer is easy, develop my tests from now on . . .” (T612)

Theme 1: Framework facilitates analysis of tests. This theme pertained to 47% of the teachers’ responses, suggesting that they believed the framework facilitated their analysis of the extent to which their tests engaged students in important mathematical processes/practices. The following comment is typical of the responses:

“Using the . . . framework helped me to evaluate whether or not test items aligned to the depth that is expected with Common Core.” (T604)

Teachers’ reactions to the overall assessment experience. One question for us has been whether the time and effort required to build teachers’ assessment knowledge relative to mathematical processes has the potential to impact their students’ learning and whether it is worth the investment of time in the courses. Responses from both preservice and in-service teachers seem to suggest that the experience has raised their awareness of the need to look at items objectively to ensure they provide opportunities for students to engage with the processes. However, responses from the in-service teachers suggested deeper thought, likely because they were reflecting on current practices and not just on future practices. Responses such as the following suggest to us that the experience is likely worth the time investment, although checking teachers’ actual classroom test practices was beyond the scope of the course or our related research endeavors.

“I think that this process has benefited me greatly in that as I analyzed each question I found myself referring back to the NCTM Process Standards for clarification and I realized very quickly that I was hindering my little problem solvers. I am on the math committee for 1st grade and we plan the unit but I create the assessments and although I use a CCSS database to draw from I elect not to use essay questions and opt for fill in the blank and multiple choice questions not realizing that I have eliminated the opportunity for the students to reason.” (T606)

“As I create my own tests or rework the tests from the textbook company, my new motto will be ‘Go below the surface, make them dig deeper.’” (T615)

“These assessment experiences confirm why I make a majority of my test and quizzes. My students are given an opportunity to show a representation of different methods and reasoning to explain work and find the answers!” (T618)

Adapting and Extending the Assessment Experience

In contrast to the undergraduate/graduate methods courses in which we use the previously described assessment experience, the graduate Current Trends course allows for more time to be spent on assessment, integrating the previously described experience with further investigations of national and international assessments (e.g., National Assessment of Educational Progress [NAEP], Trends in International Mathematics and Science Study [TIMSS], and Programme for International Student Assessment [PISA]). Table 4 highlights the topics for the first 8 weeks of the course, illustrating how assessment is related to issues of curriculum reform; readings for the week related to assessment are included in Appendix A.

Teachers’ Test Analysis and Modifications

To gauge the effectiveness of the framework to guide teachers’ analysis and modification of their own tests, we reviewed the responses to the assignment in Figure 4 from 14 teachers who provided consent. Again, our focus was not on right/wrong coding but on their overall analysis and reflection about their findings. Teachers’ work on the assignment generally reflected appropriate coding according to the framework and item modifications that were targeted to enhance the focus on processes the teachers believed were underrepresented on each test. Per step 2 of the guidelines, teachers summarized their test coding results. Although they analyzed tests of their own choosing, 13 of the 14 teachers generally found results similar to findings from our previous research using the framework (Hunsader et al., 2013, 2014). That is, tests accompanying published curricula often provide limited opportunities for students to engage with mathematical processes, with
Table 4
Topics Related to Curriculum Reform and Assessment in the Current Trends Course

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Current issues in mathematics education</td>
</tr>
<tr>
<td>2</td>
<td>Historical perspectives on curriculum up to the era of NCTM Standards</td>
</tr>
<tr>
<td>3</td>
<td>Recent national recommendations and reports, from the 1989 NCTM Standards to the 2000 Principles and Standards</td>
</tr>
<tr>
<td>4</td>
<td>Common Core State Standards and relation to Florida standards</td>
</tr>
<tr>
<td>5</td>
<td>Classroom assessments with a focus on the NCTM mathematical process standards, introduction to MPAC framework, and group and whole-class coding of sample tests</td>
</tr>
<tr>
<td>6</td>
<td>Focus on the mathematical practices</td>
</tr>
<tr>
<td>7</td>
<td>National and international assessments, including NAEP, TIMSS, and PISA</td>
</tr>
<tr>
<td>8</td>
<td>PARCC and Smarter Balanced assessments and relation to Florida State Assessments</td>
</tr>
</tbody>
</table>

Note: Bold and italics are used to indicate weeks with a particular focus on assessment. Because 4 weeks of the course are spent on assessment issues, the independent assignment accounts for about 30% of the overall course grade. Figure 4 provides the assignment guidelines along with the scoring rubric.

Reasoning and proof typically the least visible process. The visibility of the other processes vary depending on the content focus of the test. The following excerpts are typical of teachers’ responses:

“I was able to get insight on what aspects of student understanding that can or can not be learned from the tests I analyzed. . . . Reasoning and Proof was completely absent from 3 out of 5 test [sic], and added up to less than 10% of the two tests that it was in. This is a hard gap to ignore. Reasoning and proof are arguably the heart of mathematics, but it is gone in assessment practices; likely because of time constraints, yet multiple questions without mathematical processes can’t be the better option. Also intriguing is the absence of mathematical communication. Assessment should include what students are thinking . . .” (T710)

“The tests analyzed showed that the mathematical processes of reasoning and proof are not integrated into assessment. The grade 6 tests should use these processes because students at this level should begin to prepare for the use of formal proofs in their mathematics. The tests provided little opportunity for mathematical communication which does not align with the mathematical standards. Connections to the real world were used but mostly with superficial examples that does [sic] not truly show the importance of mathematics in a real world context. Overall there were a lot of graphics used but not many needed interpretation to solve the problem. Many of the graphics used were just tables that contained the information of the problem. This was just a way to represent the information but it was not necessary to interpret the graphic. Furthermore, translation of representational forms was rarely seen in these assessments.” (T709)

Step 3 of the assignment challenges teachers to use their analysis to suggest modifications for 10 test items, then describe the focus of the modifications and how the modifications would change their results from step 2. Teachers had considerable latitude in determining the focus of their modifications. As the following excerpt indicates, some chose to modify items across a range of processes:

“I tried to include all the processes in different test items that I modified. Introducing more of these processes in the assessments would allow students to present or illustrate what they are thinking and why they thought their responses were correct, in lieu of one final answer. Showing the process of solution communicates to the teacher where the student may have difficulty with a particular topic or where the student may have made a small error that would affect their final answer. Including more real-world connections in the test items would make the items more relatable to students, while allowing them to think analytically: what is the item asking of me, what would I do if I were in this situation?” (T703)
As the following excerpts indicate, some teachers focused on modifications related to a particular process they perceived as neglected:

“The process that seemed to be the most neglected in items for the five tests analyzed is Reasoning and Proof. Since this process was lacking the most compared to the other process, the modifications of the assessment items will attempt to correct the minimal evidence of assessment of Reasoning and Proof. A possible way to modify the items in order to address the disparity of this process is to require students to explain their response.” (T707)

“Most of the unit assessments are lacking in the area of student reasoning and proof, and communication. These tasks are critical as they allow for a greater insight into student understanding.” (T714)

In step 5 of the assignment, teachers consider the results of their investigation and discuss the relationship between their classroom tests and the expectations of national assessments related to the Common Core, even though Florida’s assessments are not associated with PARCC or Smarter Balanced. All teachers indicated the assessment experience enhanced their awareness of the need to be more intentional about their classroom assessment.
decisions; they recognized the need to ensure that processes reflected on high-stakes assessments, either state or national, are integrated into their classroom assessments. The following excerpts are representative of teachers’ comments:

“According to the PARCC website, the items on the PARCC assessment are classified into three separate tasks [sic] types. Type 1 assesses understanding and application, type 2 requires students to justify their work and show reasoning, and type 3 has items written in real-world context (PARCC). These task types are aligned with the items that were being investigated in this assignment and should therefore be adhered to when constructing assessments of my own.” (T703)

“This investigation provided me with awareness that reliance on the assessments that accompany textbooks may not provide adequate insight into a student’s ability to display the mathematical processes outlined by NCTM Process Standards, and CCSS Standards for Mathematical Process. . . . With the new national assessments related to Common Core (e.g., PARCC, Smarter Balanced, . . .), students will be required to show competency in the process standards.” (T707)

In fact, one teacher developed her own test by modifying items from her curriculum materials to address more of the processes. She noted:

“By doing so I created a test that was more diverse and had a variety of mathematical processes. The students overall did better compared to previous chapter tests and I, as a teacher was able to assess them more accurately.” (T702)

In summary, teachers’ work on the assignment suggests they found limited evidence of opportunities for students to engage with the processes and practices in the tests they analyzed. Using the results of their analysis, teachers demonstrated their ability to make targeted modifications in an effort to better balance students’ engagement in the processes within a single classroom test. Finally, their responses to step 5 suggest an awareness of potential disconnects between the inclusion of processes and practices on the tests they analyzed and what is expected of students in new state/national assessments. We believe our goal, to enhance teachers’ ability to be critical consumers of assessments, was realized in this case.

**Discussion and Limitations**

Our intention for the assessment experience was to provide teachers with an objective means to analyze classroom tests in relation to opportunities for students to engage with the mathematical processes, which we believe are at the heart of mathematics and of the Common Core, and to help them become critical consumers of the tests they often use as part of their assessment practice. Teachers’ reflections have corroborated our belief that teachers with all levels of experience need an awareness check, particularly with respect to the mathematical processes and practices embedded in tests. The teachers in our courses perceived the benefit of this experience for their future practice, whether we allocated a relatively small proportion of class time (1–2 weeks) or a longer period of time (up to 4 weeks) to assessment issues. Based on our prior work in analyzing tests across publishers and grade levels (Hunsader et al., 2013, 2014), we were not surprised by how infrequently and inconsistently the tests teachers analyzed engaged students in mathematical processes/practices. However, many teachers were shocked to realize that the tests they have been using miss many opportunities to elicit the kind of mathematical thinking they value and want to assess.

We continue to refine and adjust the assessment experience to ensure that it achieves our stated purpose. For instance, we have recently realized that we should make some changes that we believe would enhance the experience for preservice undergraduate teachers. We initially provided a single set of items for these teachers to both analyze and modify. However, we found that they struggled to modify items we had developed that already incorporated multiple processes/practices, contributing to a misconception that every item should engage students in every process. So, in the future, we plan to provide separate sets of items for analysis and for modification. Items provided for analysis will include a range of opportunities to engage students in mathematical processes/practices; in contrast, items provided for modification will reflect important content but be written with limited opportunities to engage in the processes/practices.

We have also learned that when having teachers strategically modify test items, it is helpful to focus their attention on the strengths of each item. Building on the strengths of existing test items is more efficient and likely more effective than creating new items from scratch or attempting to improve weak ones. We have also found that teachers benefit from actually answering both the original test items and their modified versions. Answering an original item helps teachers focus on what the item is actually
asking of students rather than what a first glance implies, particularly for items that have multiple correct answers. Answering modified versions helps teachers consider whether the complexity and length of response required by their modifications are realistic within the context of a unit test and/or for the intended grade level. For guidance on the process of modifying existing items, see Hunsader, Thompson, and Zorin (2014).

One limitation of our framework is that it does not include a criterion for problem solving. As indicated by NCTM (1989), true problem solving engages students in tasks different from typical exercises found in a textbook or as part of classroom instruction. Because preservice teachers typically have limited knowledge of the types of problems to which students have been exposed, it can be difficult for them to determine whether a task is simply an exercise or involves more non-routine solution approaches. However, in-service teachers are in a better position to determine whether a task is similar to ones their students have previously seen or solved. Hence, in our future work with in-service teachers we plan to add a criterion for problem solving to differentiate simple exercises from non-routine problems that cannot be solved by applying a procedure, formula, or algorithm known to the student.

**Direction for the Future**

Now, more than ever, teachers need to be equipped to analyze classroom tests for how items engage students in mathematical processes and practices (CCSSO, 2010; NCTM, 2000). McGehee & Griffith (2001) analyzed assessment trends across many U. S. states and found that high-stakes tests act as agents of change in curriculum design and classroom assessments. In this specific case, high-stakes tests have the potential to make a positive difference in student learning by focusing classroom instruction and assessment on important mathematical processes and practices. The two major consortia developing high-stakes assessments for the CCSSM, namely PARCC and Smarter Balanced, have woven mathematical processes and practices into their task-development guidelines. Students need opportunities to engage regularly in the kind of thinking espoused in the NCTM Process Standards and the CCSS Standards for Mathematical Practice during both instruction and assessment. Although teachers have no control over the content or processes/practices included in high-stakes tests, they can control the nature of the tests they use in their classrooms. We believe the assessment experiences outlined in this article will support teachers’ ability to analyze and modify their classroom tests. Tests that include important processes and practices support the kinds of thinking that should prepare students for success in the classroom and in their future personal and professional lives.

Whether using our framework or a different set of criteria as the basis of test analysis, we believe there is a need to increase the level of attention in teacher preparation and professional development to support teachers’ ability to be critical consumers of their classroom tests. Providing an objective tool to evaluate classroom tests and the support necessary to analyze and modify test items is a means to build teachers’ assessment skills. Classroom teachers, mathematics teacher educators, math coaches, and those who design and lead professional development experiences for teachers could use the structure of our framework and edit, add, or substitute criteria to suit their learning and assessment goals and reflect the mathematical processes and practices they value. Such knowledge could be integrated with other issues in test development (e.g., content alignment, learning objectives, the writing of multiple-choice items) that are often a part of more general measurement and assessment courses/experiences.

Our framework, with its focus on NCTM’s mathematical processes, was developed prior to the release of the Common Core Standards and its focus on mathematical practices. We do not believe our framework is incompatible with the Standards for Mathematical Practice. A comparative analysis of these practices and our framework, using the work of Koestler, Fellen, Bieda, and Otten (2013) as a guide, reveals that elements of each of the practices align with one or more of our framework criteria (see Appendix B). We believe that our criteria allow for the analysis of mathematical processes that most educators would agree are important for students’ conceptual development. We believe the complex nature of the practices and the fact that they are much less discrete than the NCTM Process Standards make it more difficult to analyze written tests for opportunities to engage with the practices without some knowledge of the textbook’s instruction, classroom instruction, or students’ responses. Analyzing based on the processes still provides insight into potential to engage with the mathematical practices.

Teachers’ reflections indicate a desire to integrate what they have learned from the experience into their own assessment practices. Future work might seek to gather longitudinal data to determine the extent to which these experiences effect real and long-term change in teachers’ assessment practices.

Some of the in-service teachers plan to share the framework with their colleagues and suggest working with school-based leaders to enact assessment experiences in their schools similar to what they experienced in our courses. Although our work has been confined to our
References


Authors

Patricia D. Hunsader, College of Education, University of South Florida Sarasota-Manatee, 8350 N. Tamiami Trail, Sarasota, FL 34243; hunsader@sar.usf.edu

Barbara Zorin, Department of Mathematics, 6650 5th Avenue N., St. Petersburg College, St. Petersburg, FL 33733; Drbzorin@gmail.com

Denisse R. Thompson, College of Education, University of South Florida, 4202 E. Fowler Ave., Tampa, FL 33620; denisse@usf.edu
Appendix A: Course Readings Related to Assessment

Undergraduate Methods Course (Preservice Elementary Teachers)

Text:

Readings:
Chapter 5 of Van de Walle text

Graduate Methods Course (In-Service Elementary or Middle Grades Teachers)

Texts:


Readings:
(completed in class using a jigsaw strategy)

For elementary teachers:


For middle-grades teachers:
Enhancing Teachers’ Assessment of Mathematical Processes


**Graduate Current Trends Course (Preservice and In-Service Elementary, Middle-Grades, or High School Teachers)**

**Texts (referenced during the assessment experience):**


**Readings (varied based on grade level of teachers):**

**Week devoted to Classroom assessments, with a focus on the process standards:**


**Process standards from the NCTM Principles and Standards (all teachers):**


**Week devoted to Mathematical Practices:**


**Week devoted to National and International Assessments, NAEP, TIMSS, PISA:**

**Skim based on appropriate grade level:**


Programme for International Student Assessment. (2009). *Take the test: Sample questions from OECD’s PISA assessments.* OECD.


**Week focused on Explorations related to PARCC, Smarter Balanced, and connections to Florida tests and classroom assessments:**


(Ref to page 73)
Appendix B: Comparison of CCSSM’s Standards for Mathematical Practice and the Criteria in the MPAC Framework

<table>
<thead>
<tr>
<th>Standards for Mathematical Practice</th>
<th>Alignment to MPAC Framework Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Make sense of problems and persevere in solving them</td>
<td>Connections, Representation: Graphics, Representation: Translation</td>
</tr>
<tr>
<td>2 Reason abstractly and quantitatively</td>
<td>Communication, Connections, Representation: Translation</td>
</tr>
<tr>
<td>3 Construct viable arguments and critique the reasoning of others</td>
<td>Reasoning and Proof, Communication</td>
</tr>
<tr>
<td>4 Model with mathematics</td>
<td>Communication, Connections, Representation: Translation, Representation: Graphics</td>
</tr>
<tr>
<td>5 Use appropriate tools strategically</td>
<td>Representation: Graphics</td>
</tr>
<tr>
<td>6 Attend to precision</td>
<td>Communication, Reasoning and Proof</td>
</tr>
<tr>
<td>7 Look for and make sense of structure</td>
<td>Connections, Representation: Graphics</td>
</tr>
<tr>
<td>8 Look for and express regularity in repeated reasoning</td>
<td>Reasoning and Proof, Connections</td>
</tr>
</tbody>
</table>
Developing a Mathematics Instructional Practice Survey: Considerations and Evidence*

Michele B. Carney  
*Boise State University*

Jonathan L. Brendefur  
*Boise State University*

Gwyneth R. Hughes  
*Boise State University*

Keith Thiede  
*Boise State University*

**As mathematics teacher educators, it is imperative that we have high-quality tools that conceptualize and operationalize mathematics instruction for large-scale examination.** We first describe existing instructional practice survey scales, including their conceptualization of practice and related validity evidence. We then present the framework and initial validity evidence for our mathematics instructional practice survey. Survey participants were in-service teachers in a statewide mandated mathematics professional development course. Statistical analyses indicate the items measure two constructs: social-constructivist and transmission-based instructional practice. Of particular interest is the result that these two constructs were negligibly correlated. This is in contrast to the generally accepted notion that social-constructivist and transmission-based instructional practices are the two polar ends of a single construct for describing instructional practice.

**Key words:** Instructional Practice; Mathematics; Survey Development; Validation

Over the past few decades, educational practice and research have shifted to be more data driven, as is evidenced by both public policy (NCLB, Race to the Top, Common Core) and research initiatives (Institute of Education Sciences & National Science Foundation, 2013). Teachers’ instructional practices in the classroom have been shown to be central to student achievement (Boaler, 2005; D’Agostino, 2000), yet studying these practices objectively is inherently difficult given the complexity of classroom instruction. Despite these difficulties, there is a need to define and measure instructional practice in order to evaluate programs and conduct quantitative educational research. But how does one decide what “instructional practice” is, and how does one measure it?

To examine practice meaningfully it must be explicitly conceptualized and then translated into an instrument that appropriately operationalizes and measures this conceptualization. Too often one of two things occurs: Instruments are developed without clear conceptualization of practice, or instruments are used in research or evaluation that do not adequately match the conceptualization of practice within the project. Either disconnect can lead to a lack of significant findings due to the measure rather than the variables under examination. The mathematics education community needs multiple measures of instructional practice that vary in conceptualization and that can be used on a large scale (e.g., self-report surveys) to examine and evaluate different aspects of instructional practice.

The goals of this article are twofold. We first examine existing self-report measures of instructional practice, discussing the conceptualization of instructional practice within each and associated validity evidence. We then conceptualize instructional practice as it relates to our own professional development model and describe why and how we developed our own instrument to operationalize those practices. Lastly, we present the results of two validity evidence studies on our instrument. This article is meant both as a tool for other researchers in thinking about implementing or developing measures of instructional practice, and a report on our own findings with regard to our specific survey instrument.

**Operationalizing Instructional Practice—An Overview**

Although self-report surveys are useful for large-scale studies, using them to measure a construct as complex as instructional practice is inherently crude. The relativistic nature of what “good” teaching means is an inherent problem in the reliability of teachers’ assessments of the quality of their own teaching (Mayer, 1999; Wubbels,
1992). However, Mayer (1999) found that teachers’ reports of frequency of specific instructional practices were relatively accurate as compared with the validity issues that arise when teachers report on the quality of instructional practice. Analysis of practice from a frequency perspective, instead of a quality perspective, may increase the agreement between teachers’ self-reported practice and observations of actual classroom practice (i.e., the validity of the measure). But it decreases the specificity of what can be detected by the measure, hence its inherent crudeness. This does not mean we should not attempt to measure instructional practice through a self-report survey. Rather, findings based on the survey data should be interpreted relative to the type of measurement that can accurately be performed and the validity evidence provided to support those findings.

Establishing Validity Evidence for Survey Measures of Instructional Practice

Mathematics teacher educators must be critical consumers of research instruments. To this end, we provide a brief explanation of how one establishes validity evidence for psychometric instruments and an overview of commonly used survey scales of instructional practice. Our purpose in this section is to provide mathematics teacher educators with a framework for evaluating validity evidence for survey scales of instructional practice, including our own.

In research we often need to measure a concept of interest. Some concepts can be directly measured, such as height. However, in educational research we often need to measure human-formulated concepts, termed constructs, such as mathematics instructional practice. Such constructs are indirectly measured through instruments, such as surveys, that typically contain a grouping of items that together measure the construct. Following appropriate statistical analyses, the individual item scores are combined to provide a scale score reflecting the degree to which an individual demonstrates or possesses a particular construct. Validity for an instrument refers to the strength of evidence that the scale score accurately represents the level an individual possesses of the construct of interest. For example, teachers’ ability to respond to student thinking in the mathematics classroom may be the construct of interest. Survey developers may build a scale around that construct with several items that, when combined into a scale score, reflect the nature of a teacher’s ability in that area. However, in addition to providing the scale items, developers should provide evidence demonstrating that a high (or low) scale score on the survey actually reflects the presence (or absence) of the construct of interest. As mathematics teacher educators, our selection of instruments should be informed by our particular research context and question(s), and we need to closely examine the validity evidence provided.

In discussing a particular instrument, researchers often refer to the validity of the instrument itself. However, current theory focuses on establishing the validity of the evidence that supports the interpretation of an instrument’s scores, specific to the context in which the validity evidence was collected (Kane, 2006). For example, if a survey has been validated by evidence from classroom observations of elementary preservice teachers, then the same validation evidence may not support interpretation of scale scores from high school mathematics teachers. The burden is on the researcher to select instruments with validity evidence supporting use of the instrument within their particular research context. We have found that the validation framework from Cook and Beckman (2006), consisting of content, response process, internal structure, and relationship to other variables (described in the paragraphs that follow), provides a thorough yet simple approach to considering validity evidence for psychometric instruments.1

Content. Does the instrument measure the depth and breadth of the construct under examination? This involves first clearly defining how the construct is conceptualized and how it will be operationalized. The developer must then provide evidence that the instrument addresses the full depth and breadth of the construct by, for example, providing a framework for the development of the items. If there is no clear evidence of content validity, it does not mean the developers did not have a clear conceptualization or framework, just that this evidence has not been explicitly provided. In that case the burden shifts to the researcher who plans to use the scale to ensure it addresses the depth and breadth of the construct as needed for their research or evaluation purpose.

Response process. Is there a match between the thought process of the respondent and the intended construct under examination? For surveys, this involves providing evidence that respondents are interpreting the survey items as intended, such as by conducting cognitive interviews (Desimone & Le Floch, 2004). A related consideration when researchers intend to use survey scales to evaluate change from pre- to post-intervention is response-shift bias (Bray, Maxwell, & Howard, 1984), which involves validity threats that arise when respondents’ understanding and interpretation of survey items shift as a result of participation in an intervention. For

---

1 Cook and Beckman (2006) also address consequential validity.
example, teachers’ understanding of a survey item related to what it means to ask students to justify their answers can change significantly as a result of professional development. This can often be addressed through a retrospective pre/post survey design that involves asking participants to recall their preintervention status postintervention (Lam & Bengo, 2003).

Internal structure. Is there an acceptable level of evidence regarding internal consistency and construct structure? This refers to evidence that the measurement error around a set of items is minimal (reliability) and that the items collectively measure a common construct (unidimensionality). Cronbach’s alpha is often used to measure the internal consistency of a set of survey items. It provides an indicator of the correlation between survey items and the likelihood of obtaining similar responses if readministered. A low Cronbach’s alpha indicates the survey items may not be accurately or consistently measuring the construct under investigation and suggests that a researcher should not compute a scale score for the survey items because they are measuring more than one concept of interest. In addition to Cronbach’s alpha, the use of statistical techniques, such as factor analysis (FA), examines the unidimensionality of the survey scale, ensuring the items measure a single construct rather than multiple constructs that are correlated with each other.

Relationship to other variables. Is there an appropriate level and direction of correlation to other variables of interest? This involves multiple aspects of validity evidence, including criterion, convergent/discriminate, concurrent, and predictive. For example, criterion validity typically refers to evidence that the instrument under examination produces scores that are correlated to a related criterion measure. For the construct of instructional practice, this could be demonstrated by providing a correlation coefficient between self-report survey scale scores of instructional practice and observations of classroom practice. Predictive validity evidence indicates the ability of one variable (independent) to predict another (dependent). For example, instructional practice is often examined in relation to student achievement measures. However, due to the multitude of factors influencing student achievement, the expected predictive between these variables is often difficult to demonstrate.

As may be evident from the description of the various elements of the validity framework, the process of creating, modifying and examining the validity evidence for survey items and scales is an intensive, iterative process that should not be undertaken lightly. Ideally, researchers will identify a scale in the literature that conceptualizes practice in a manner that meets their needs. Below we provide a brief description of four instructional practice survey measures and their associated validity evidence.

Construct Validity Evidence for Existing Instructional Practice Survey Instruments

There are multiple measures of instructional practice available for researchers and practitioners to use in research and evaluation. Our review focused on identifying self-report instructional practice survey scales with the following characteristics: (1) they are described in published, peer-reviewed articles focused on presenting validity and reliability evidence, (2) the sample used in-service (versus preservice) teachers, and (3) the response categories focused on frequency of occurrence of particular practices (versus level of agreement/beliefs). Due to the limited number of peer-reviewed articles identified, we also included reports of survey development and validation that were not peer reviewed but provided sufficiently detailed analysis. We identified four survey scales that fit these criteria. In the following section we briefly examine these survey scales through the lens of content, response process, internal structure and relationship to other variables (Cook & Beckman, 2006). Table 1 also provides some general information about each of the scales. Please keep in mind our analysis is based on the information we found in the documents we reviewed. Because it is difficult to publish the full breadth of survey validation evidence, it may be that additional research was conducted but not reported or was missed in our review process.

Horizon-Reform. This four-item scale assesses the use of “reform-oriented teaching practices” and is a part of the much larger Horizon Research 2012 National Survey of Science and Mathematics Education (Banilower et al., 2013). The construct of reform-oriented teaching practices is conceptualized as students’ use and explanation of multiple approaches to solving a mathematics task. The survey report indicates response process was examined through cognitive interviews that were conducted for the entire survey (Banilower et al., 2013, p. 4). Internal structure evidence was provided through examination of the unidimensionality of the items in the scale through factor analysis and providing a Cronbach’s alpha (Banilower et al., 2013, pp. E–16). Due to the overall breadth of the National Survey of Science and Mathematics Education, there is little content validity evidence provided for this particular scale, and relationship to other variables was not addressed in the report.

TIMSS-Engagement. This four- or six-item scale for eighth- or fourth-grade teachers, respectively, assesses the use of “instruction to engage students in learning”
Table 1: General Survey Information on a Sample of Published Self-Report Instructional Practice Survey Measures for Practicing Teachers

<table>
<thead>
<tr>
<th>Survey</th>
<th>Original purpose</th>
<th>Sample</th>
<th>Constructs (No. of items)</th>
<th>Internal structure</th>
<th>Sample item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizon-Reform</td>
<td>Evaluate mathematics instructional practice from a national sample and examine relationship to equity variables</td>
<td>National samples of U.S. practicing elementary, middle, and high school teachers</td>
<td>Reform-oriented teaching practices (4)</td>
<td>α = .77 and CFA</td>
<td>Have students explain and justify their method for solving a problem.</td>
</tr>
<tr>
<td>TIMMS-Engagement</td>
<td>Provide comparison of mathematics instructional practice across countries and relationship to achievement</td>
<td>International samples of practicing 4th- and 8th-grade teachers</td>
<td>Engage students in learning (4th: 6 items) (8th: 4 items)</td>
<td>4th: α = .4 to .83 8th: α = .18 to .76 and CFA</td>
<td>Praise students for good effort.</td>
</tr>
<tr>
<td>Swan-Practices</td>
<td>Evaluate the influences of a mathematics professional development project.</td>
<td>184 FE teachers (two groups)</td>
<td>Instructional practices (25 items)</td>
<td>α = .85 No CFA</td>
<td>I am surprised by ideas that come up in a lesson.</td>
</tr>
<tr>
<td>Ross-Commitment</td>
<td>Evaluate the commitment to implementation of standards-based mathematics teaching</td>
<td>2687 K–8 teachers in U.S. (two groups)</td>
<td>Commitment to standards-based mathematics teaching (20)</td>
<td>α = .81 to .88 No CFA</td>
<td>I tend to integrate multiple strands of mathematics within a single unit.</td>
</tr>
</tbody>
</table>

1 Used reverse coding of particular items (i.e., viewed instructional practice as a continuum).
CFA = Confirmatory factor analysis

and is part of the International TIMSS 2011 Context Questionnaire administered to teachers whose students participate in the Trends in International Mathematics and Science Study (TIMSS) assessment (Martin & Mullis, 2012). “Instruction to engage students in learning” is conceptualized as encouraging, engaging, and questioning students during various aspects of mathematics instruction. Internal structure evidence was provided through examination of the unidimensionality of the items within the scale using factor analysis. From an internal-consistency perspective, the scale analyses offer an interesting perspective on the importance of considering the context and sample from which the validity evidence is drawn. The six-item scale had an alpha that ranged from .40 for teacher respondents in Azerbaijan to .83 in Chinese Taipei. The range in the four-item scale was considerably larger, .18 to .76. These large discrepancies in scale reliability across countries emphasize the need to consider the sample and context from which the internal structure evidence is drawn when determining whether or not the scale will hold together in a different research or evaluation context. Relationship to other variables was examined through correlational analysis, and essentially no consistent relationship between student responses and students’ performance on the TIMSS were found. Content and response process were not specifically addressed within the report.

Swan-Practices. This 25-item scale assesses “teacher-centered practices” arising from a transmission-oriented belief system or “student-centered practices” arising from a constructivist-oriented one (Swan, 2006). “Teacher-centered practice” is conceptualized as the teacher transmitting knowledge to students, while “student-centered practice” is conceptualized as taking students’ individual processes for knowledge-building into account through instruction designed to be flexible to student needs. This questionnaire was developed in conjunction with a beliefs questionnaire to assess changes as a result of professional development provided by Further Education colleges in England. Content validity evidence is provided through the framing of practice as arising from transmission, discovery, or constructivist beliefs about mathematics teaching and learning (Ernest, 1989). Response process
evidence was provided through the use of instructional practice vignettes, Cronbach's alpha was reported as a mean of evidence for internal structure, and observations of classroom practice and students' descriptions of teachers' practices were provided as criterion validity evidence for relationship to other variables.

**Ross-Commitment.** This 20-item scale assesses elementary teachers’ “commitment to mathematics education reform” (Ross, McDougall, Hogaboam-Gray, & LeSage, 2003). It conceptualizes this commitment a broad array of topics, from giving students open-ended problems to the teacher acting as a facilitator rather than a transmitter of knowledge. The authors provide content validity evidence through a clear framework for the development of the survey items based on the NCTM standards (2000) and examination of items by a panel of elementary mathematics specialists. Internal structure evidence is given in the form of Cronbach's alpha, and evidence of relationship to other variables is provided through examination of the relationship between the scale score and school-level means for a test of sixth grade mathematics achievement. Response process is not specifically addressed in the article.

**Operationalizing Instructional Practice—the DMT Framework**

We chose to undertake the development and validation process for our instrument because we wanted the operationalization of instruction to closely match the Developing Mathematical Thinking (DMT) framework for instructional practice so we could measure the influence of our professional development on a large scale. The existing measures either did not closely match on conceptualization of practice, or in some cases they conceptualized practice as a continuum (e.g., Swan-Practices and Ross-Commitment), which could be measured by a single scale. We found this problematic based on our experience in classrooms observing teachers’ instructional practice. It is important to note we did not undertake the development of our own survey scale lightly. The development of a survey scale and process for gathering and presenting validity evidence is extremely complicated. The decision to develop one’s own scale should be considered in light of the time and resources needed to create a high-quality scale with regard to the various validity elements.

For our purposes, the Swan-Practices scales, which use learning theories as a lens for thinking about instructional practice, provided a basis for conceptualizing our scales. We chose to develop our own items rather than use the Swan-Practices scales based on our fundamentally different conceptualization of instructional practice. In particular, we did not want to make the assumption that what are commonly referred to as teacher-centered and student-centered practices lie at opposite ends of a single construct’s continuum. We instead view student-centered and teacher-centered practices as related but independent subconstructs within instructional practice. The next sections provide content validity evidence by describing the conceptual domains of the DMT framework and the development of the initial survey framework and items.

**Developing Mathematical Thinking (DMT) Theoretical Framework**

The DMT framework is built upon social and cognitive learning theories, which hold that students need to learn mathematics by constructing knowledge through meaningful classroom activities and discussions. The teacher’s role in the classroom is to facilitate student learning through the meaningful selection of mathematical tasks and high-quality classroom discussion designed to build connections between students’ informal knowledge and the formal knowledge of mathematics that has developed over time (Carpenter, Fennema, Franke, Levi, & Empson, 1999; Gravemeijer & van Galen, 2003; Hiebert, 1997). Our framework focuses on five classroom instructional practices that develop students’ mathematical understanding: (1) taking students’ ideas seriously, (2) pressing students conceptually, (3) encouraging multiple strategies and models, (4) addressing misconceptions, and (5) focusing on the structure of the mathematics (Brendefur, Carney, Hughes, & Strother, forthcoming; Carney, Brendefur, Thiede, Hughes, & Sutton, 2014). A brief description of each of the domains follows.

**Taking students’ ideas seriously (TSIS).** TSIS involves valuing and building upon students’ intuitive understanding of mathematical concepts (Carpenter & Lehrer, 1999; Hiebert, 1997; Romberg & Kaput, 1999). For example, when students solve an unfamiliar yet meaningful math problem, they draw on their prior knowledge and experience. Their solution strategies and notations may seem inefficient or informal to an observer, but by eliciting and valuing students’ initial solution strategies, teachers can connect student thinking to more efficient and abstract methods (Freudenthal, 1973, 1991; Gravemeijer & van Galen, 2003; Treffers, 1987).

**Pressing students conceptually (PSC).** PSC focuses on building connections between mathematical strategies and models and progressively formalizing those ideas and methods for solving problems (Carpenter & Lehrer, 1999; Forman, 2003; National Research Council, 2001; Siegler, 2003). For example, once students have had the chance
to work on their own solution methods, teachers press them to connect and compare methods, generalize them to new situations, and relate them to formal mathematical terms and conventions. It is through this process of connection and generalization that students move from their own informal methods to more formal and efficient strategies (Carpenter & Lehrer, 1999; Gravemeijer & van Galen, 2003).

**Encouraging multiple strategies and models (EMMS).** EMMS involves developing students’ understanding of various models and approaches to solving problems (Dolk & Fosnot, 2006; National Council of Teachers of Mathematics, 2000; Romberg & Kaput, 1999). When students generate, evaluate, and utilize different mathematical strategies and models, they recognize there are many ways to solve problems and represent solutions (Bruner, 1964). In addition, different strategies and models highlight different aspects of the mathematics, and thus examining the same problem through different lenses deepens students’ overall understanding of the topic.

**Addressing misconceptions (AM).** AM involves using students’ mistakes and misconceptions as valuable tools to build mathematical understanding (Borasi, 1987, 1994; Bray, 2013; Gooding & Stacey, 1992). Making mistakes and learning from them is an integral part of doing mathematics at any level. But mistakes often recur even after teachers demonstrate a correct procedure because they stem from deeper mathematical misconceptions. By being aware of why and how misconceptions develop and taking the time to address misconceptions through models and discussion, teachers can move students to a deeper level of understanding that precludes such mistakes. Additionally, mistakes can be opportunities for students to engage in justification, evaluation, and inquiry (Borasi, 1987).

**Focusing on the structure of the mathematics (FSM).** FSM involves facilitating students’ understanding of fundamental, or structural, mathematical concepts (e.g., decomposing and composing, units and unitizing, equivalence). Many teachers and their students see mathematics as a series of procedures and definitions that build in complexity throughout the K–12 curriculum. But certain fundamental ideas or “structural components” appear continually throughout mathematics, whether one is looking at 2nd or 11th grade. Focusing on these structures allows students to build understanding of and establish connections between these fundamental concepts and the particular topic being studied (National Council of Teachers of Mathematics, 2000; National Governors Association & Chief Council of State School Officers, 2011). When instruction does not focus on the structure of mathematics, students often rely on memorized tricks or formulas and have difficulty solving complex problems or applying mathematics to new situations.

**Social-Constructivist and Transmission-Based Instructional Practices**

In addition to examining the breadth of the mathematics instructional practice construct, it is important that surveys address the various perspectives within that domain. Teachers’ instructional practice is often polarized in the mathematics instructional practice survey literature as either teacher-centered versus student-centered (e.g., Swan, 2006) or given other similar labels. While these labels simplify interpretation of the constructs under examination, they place the practices associated with each construct on opposite ends of a continuum without recognizing that teachers can effectively use both instructional approaches, depending on the situation or topic. For example, in a data unit, a teacher might utilize an exploratory activity and classroom discussion to generate ideas about how to represent and analyze the heights of students in the class. Later in the unit, the same teacher might directly instruct students about how to create a box and whisker plot, in line with mathematical conventions and norms. This direct instruction might then be followed by small-group discussions comparing different formal representations. Thus, rather than being at odds, transmission-based and social-constructivist learning may in fact be used simultaneously, and it would in fact be worrying if only one or the other were in constant practice. In addition, some labels themselves, for example “student-centered,” tend to legitimize particular instructional practices, while others, such as “teacher-centered,” are deprecating. If one of the purposes of measuring instructional practice is to determine the relationship of instructional practice to important variables of interest, it is necessary to use labels that accurately describe the trait under examination while also allowing for ease of interpretation of the meaning.

Based on our professional development, we wanted a survey that would allow teachers to realistically describe actual classroom practice without obviously “penalizing” a certain type of practice. We borrowed from Swan’s (2006) framing of instructional practice as arising from different beliefs about how students learn. We utilized transmission-based (Cobb, 1988; Stipek, Givvin, Salmon, & MacGyvers, 2001) and social-constructivist (Cobb & Yackel, 1996) learning theory constructs to identify instructional practices and terminology that could arise in or describe a classroom based on each of these learning theories. We used those terms to frame instruction typically associated with teacher- and student-centered instructional practices, respectively. We recognize that these terms are typically associated with learning theories...
and can manifest in multiple ways with regard to instructional practice. The DMT components provide a specific framework for instruction built upon these theories.

The next section describes the survey development process, addressing evidence of content validity. This is followed by the methods and results for Study 1, which examines the evidence of internal structure and capacity for the survey instrument to evaluate change in pedagogy by analyzing data from participants who took a 3-credit professional development course. Last, the methods and results for Study 2 examine the relationship of the survey constructs to observations of instructional practice utilizing data from participants involved in a 3-year professional development grant.

Survey Item Development

Our survey framework utilized three perspectives of instructional practice to ensure we measured the depth and breadth of the domain. The first perspective involved the use of the five domains within the DMT framework previously described (see the column headings in Table 2). This was followed by examining each DMT domain through the perspective of student practices, teacher practices, and classroom tasks and activities to address the depth within each domain (see the row headings in Table 2). Last, the transmission-based (T) and social-constructivist (S-C) learning theory perspectives were operationalized through survey items written to address each perspective. Table 2 provides an overview of the theoretical framework used in the initial development of survey items. We anticipated that the constructs of social-constructivist and transmission-based practices would emerge from the data.

Initially, one to three items were placed in each of the 30 cells of the framework. A total of 74 items were constructed. For example, in the domain of taking students’ ideas seriously, a social-constructivist item with a student perspective was, “Students are encouraged to discuss their mathematical ideas in pairs, small-group, and/or whole-class discussions.” A transmission item with a teacher perspective was, “I demonstrate for the class the correct way to use a particular procedure or model before they start solving problems.”

To help establish the content validity of the instructional practice scale, a panel of six university-level mathematics education professors and professional developers analyzed the items. Any items deemed to inaccurately measure the construct domain were revised or removed. Once the initial development and review process was completed, the process took on a cyclical nature: survey item administration, analysis of the data in relation to variables of interest and psychometric properties, and finally revision and review. This process occurred three times and led to the refinement of the initial set of 74 items to 30 items, one addressing each cell within the framework.

Study 1

Study 1 focused on providing validity evidence related to internal structure and relationship to other variables. The research questions guiding Study 1 were:

1. Which items most strongly correlate to the constructs of social-constructivist and transmission-based instructional practice?

Table 2

Theoretical Framework to Address the Depth and Breadth of Classroom Mathematics Instructional Practice.

<table>
<thead>
<tr>
<th>Perspective</th>
<th>Take students’ ideas seriously</th>
<th>Encourage multiple models and strategies</th>
<th>Press students conceptually</th>
<th>Address misconceptions</th>
<th>Focus on the structure of mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>T item(s)</td>
<td>T item(s)</td>
<td>T item(s)</td>
<td>T item(s)</td>
<td>T item(s)</td>
</tr>
<tr>
<td></td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
</tr>
<tr>
<td>Teacher</td>
<td>T item(s)</td>
<td>T item(s)</td>
<td>T item(s)</td>
<td>T item(s)</td>
<td>T item(s)</td>
</tr>
<tr>
<td></td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
</tr>
<tr>
<td>Tasks &amp; activities</td>
<td>T item(s)</td>
<td>T item(s)</td>
<td>T item(s)</td>
<td>T item(s)</td>
<td>T item(s)</td>
</tr>
<tr>
<td></td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
<td>S-C item(s)</td>
</tr>
</tbody>
</table>
2. Using a separate sample, do the survey items correspond to their intended constructs?

3. What is the relationship between the constructs of social-constructivist and transmission-based instructional practice?

4. Do the survey scale scores for the two constructs have the capacity to evaluate change in intended practice as a result of professional development?

Data Source

The data for the present study came from participants enrolled in a three-credit, 45-hour mandated mathematics professional development course for kindergarten–12th grade teachers and administrators called Mathematical Thinking for Instruction (MTI). The course focuses on enhancing teachers’ knowledge of mathematics, their understanding of how students best learn mathematics, and their ability to teach mathematics effectively. MTI instructors utilize the DMT framework—(1) taking students’ ideas seriously, (2) encouraging multiple strategies and models, (3) pressing students conceptually, (4) focusing on the structure of the mathematics, and (5) addressing misconceptions, both implicitly and explicitly throughout the course. The MTI course uses number and algebra topics as the basis for understanding and applying the DMT framework. The professional development activities utilize participants’ thinking, strategies, and models as the basis for discussion, thus modeling “taking students’ ideas seriously” through both intensive tasks such as examining dividing fractions (e.g., Hughes, Brendefur, & Carney, 2015) or smaller investigations such as exploring multiplication facts through arrays. The ensuing group discussion is modeled around the DMT framework by pressing connections between participants’ models and strategies and explicating the structure of the mathematics and potential misconceptions. This is often followed by explicit discussion of applications of the content and DMT framework to the K–12 classroom.

Survey Administration

Participants received a link to the instructional practice survey via email following course completion during the spring and summer of 2013. While the majority of course participants were teachers actively instructing in mathematics, the group also included (1) administrators, such as principals, district office personnel, and superintendents, (2) special education teachers who may or may not teach mathematics on a regular basis, and (3) all K–8 certified personnel, which included many middle school teachers of nonmathematics content. Individuals who did not teach mathematics on a regular basis were eliminated from the analysis, as indicated in the last column of Table 3.

Participants were asked to respond to the survey items following the completion of the MTI course (referred to as after) and through retrospective analysis of their practice prior to participation in the course (referred to as prior). Retrospective analysis was utilized because participants were likely to have a different understanding of the survey item wording and constructs following participation in the MTI course (Aiken & West, 1990; Lam & Bengo, 2003). Our approach of reporting postcourse and retrospective precourse instructional practice is the more conservative approach when compared with other forms of retrospective analysis that specifically measure change in practice (Lam & Bengo, 2003).

We had an overall survey response rate of 85.8%, with a total of 798 applicable course participants who taught mathematics on a regular basis completing the course and end-of-course survey during the spring and summer of 2013. For each participant, the data set contained 30 survey items with responses related to their instruction prior to the MTI course and the same 30 items with responses related to their instruction after the MTI course. The survey response scale ranged from 1 = Never to 7 = Daily (1 = Never, 2 = 2–3 times per year, 3 = Once per month, 4 = 2–3 times per month, 5 = Once a week, 6 = 2–3 times per week, 7 = Daily). Given that the MTI course could span from 1 to 10 weeks and our survey scale spanned frequency of practice from daily to 2–3 times per year, what was likely captured in participants’ responses after the MTI course was their intended, instead of actual, frequency of practice.

The survey items were grouped according to three perspectives: student, teacher, or task/activities. This design was intentional because we did not expect these aspects of the framework to emerge as constructs (i.e., latent variables) in the initial analysis. Within a survey page, the

<table>
<thead>
<tr>
<th>Course level</th>
<th>Course participants</th>
<th>Survey participants</th>
<th>Applicable participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>K–3</td>
<td>933</td>
<td>828</td>
<td>497</td>
</tr>
<tr>
<td>4–8</td>
<td>550</td>
<td>429</td>
<td>229</td>
</tr>
<tr>
<td>6–12</td>
<td>141</td>
<td>136</td>
<td>72</td>
</tr>
<tr>
<td>Total</td>
<td>1624</td>
<td>1393</td>
<td>798</td>
</tr>
</tbody>
</table>
item stem specific to that perspective was placed at the top of the page with all related survey items below. The item order was randomized for each survey participant. The purpose of mixing construct items both within and across the pages of the survey was to ensure the constructs still held together when intentionally mixed. An example from the tasks/activities perspective page of the survey is provided in Appendix A.

**Survey Refinement**

As described previously, our survey development process resulted in 30 items: 15 designed to address social-constructivist practices and 15 designed to address transmission-based practices. The reality of survey development is that despite our best efforts, these items may or may not be good indicators of these two constructs. Part of the survey development process involves conducting statistical tests (e.g., exploratory factor analysis and confirmatory factor analysis) to determine which items are most highly correlated with (i.e., best measure) the constructs under examination. Once it has been verified that items cleanly measure a construct, it is more appropriate to look at a total or average of those items as a representation (or operationalization) of that construct, often referred to as a scale score. Keep in mind that this numeric score still has a certain amount of measurement noise or variability (i.e., it does not perfectly measure the construct), but is a clearer representation of a unidimensional construct than if all the original items had been retained to create the scale score.

The purpose of the following analyses is to examine the internal structure of a survey of mathematics instructional practice. Based on our a priori theoretical framework, we anticipated latent variables around the constructs of socialist-constructivist and transmission-based learning theories. Confirmatory factor analysis (CFA) was utilized to examine the validity of the hypothesized factor structure. However, exploratory factor analysis (EFA) was initially utilized as a starting point to identify mis-fit items or potential subconstructs within the anticipated latent variables. The MTI course data set \( n = 798 \) was randomly divided into two datasets so the two stages of analysis could be performed on independent samples. Following examination of the factor structure, the entire data set was analyzed to determine whether significant changes could be detected in the latent variables from the prior vs. after perspectives.

**Results**

**Exploratory factor analysis (EFA) and item analysis.**

Exploratory factor analysis was used to determine which survey items did the best job of measuring (or were most highly correlated with) our constructs of social-constructivist and transmission-based practice. The technical write-up, factor loadings, and summary statistics are presented in Appendix B. Eight of the 15 items were identified as cleanly measuring the construct of social-constructivist practice, and 6 of the other 15 items were identified as cleanly measuring the construct of transmission-based instructional practice. The items and associated constructs are presented in Table 4, and an example of these items in a survey format is provided in Appendix C.

In addition to the exploratory factor analysis of item correlations, means and standard deviations were computed. These are provided in Appendix D for prior (Table D1) and after (Table D2) MTI course participation. The items within each identified construct significantly correlated to one another. In other words, social-constructivist items correlated significantly with one another, and transmission items correlated significantly with one another. In addition, internal consistency was examined for each scale and was found to be good to excellent for all scales: social-constructivist \( \alpha = .91 \); transmission-based \( \alpha = .90 \). This suggests that the scale and was found to be good to excellent for all scales; social-constructivist \( \alpha = .91 \); transmission-based \( \alpha = .90 \).

**Confirmatory factor analysis (CFA).** The purpose of confirmatory factor analysis is to determine whether the sample data do a reasonable job of matching a hypothesized model. The major components of our hypothesized model included (a) two latent variables or constructs associated with social-constructivist (soccon) and transmission-based (trans) practice, and that (b) each survey item would correlate to the latent variable it was designed to measure and not correlate to the other latent variable. Our analyses indicated that the sample data do a reasonable job of matching a hypothesized model, confirming the results of our EFA (detailed CFA findings in Appendix E). Second, the construct of social-constructivist instructional practice was for the most part uncorrelated to transmission-based instructional practice. The lack of a moderate to strong negative correlation between the two constructs indicates our survey participants view these as distinct and unrelated constructs within mathematics instructional practice, rather than as two ends of a single instructional spectrum.

**Capacity to evaluate change.** Following the EFA and CFA analyses, the entire data set was combined and evaluated for the capacity of the scale scores for social-constructivist and transmission-based practice to capture change as a result of professional development. While a lack of change from prior to after would not invalidate the survey’s capacity to capture change—the lack of change could be due to the professional development itself—the
demonstrated ability to capture change from a professional development course that has already successfully demonstrated significant changes in teachers’ knowledge, beliefs, and self-efficacy (Carney, Brendefur, Thiede, et al., 2014) assists in the initial validation efforts. Scale scores were constructed by finding the mean of the eight items for social-constructivist and six items for transmission-based practices, eliminating those individuals missing more than one item score within a scale.

Appendix F provides histograms for each of the variables from the perspective of prior and after MTI course participation. The histogram for social-constructivist practice after MTI course participation indicates a potential issue with the frequency of practice scale; there is a ceiling effect on reporting particular types of practice.

A Wilcoxon signed-rank test showed that participants’ self-reported, retrospective analysis of their instructional practice indicated significant changes across both constructs; social-constructivist \((Z = 22.718, p < .001)\) and transmission \((Z = 20.072, p < .001)\). The scores can be interpreted within the original scale metric; 1 = Never, 2 = 2–3 times per year, 3 = Once per month, 4 = 2–3 times per month, 5 = Once a week, 6 = 2–3 times per week, 7 = Daily. The median for social-constructivist practice was 4.6 \((n = 847)\) for prior and 6.6 \((n = 784)\) for after MTI course participation. On the original metric, this would indicate that prior to the MTI course, teachers reported engaging in social-constructivist practices less than once a week; following the MTI course, they indicated a shift in social-constructivist practice to more than 2–3 times per week. The median for transmission-based instructional practice demonstrated similar changes to the social-constructivist variable but in the opposite direction, with 5.7 \((n = 857)\) prior to and 3.6 \((n = 772)\) after the MTI course. On the original metric, this indicates a shift from conducting transmission-based practice more than once a week before the course to less than 2–3 times per month after the course.

### Study 2

Study 2 focused on providing validity evidence related to the relationship between variables of interest: What is the relationship between teachers’ self-reported instructional practice survey scores (both social-constructivist

---

**Table 4**

**Survey Items and Associated Constructs Retained for CFA**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Survey items (DMT framework element)</th>
<th>Item label for CFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social-constructivist items</td>
<td>Facilitate small-group or whole-class discussion on student thinking (TSIS_T)</td>
<td>SC1</td>
</tr>
<tr>
<td></td>
<td>Are based on their potential to encourage discussions of students’ mathematical ideas (TSIS_A)</td>
<td>SC2</td>
</tr>
<tr>
<td></td>
<td>Emphasize the use of multiple models for recording and communicating student thinking (EMRS_T)</td>
<td>SC3</td>
</tr>
<tr>
<td></td>
<td>Solve problems that allow for several different approaches (EMRS_S)</td>
<td>SC4</td>
</tr>
<tr>
<td></td>
<td>Facilitate discussion about underlying mathematical concepts (e.g., composing or decomposing number; FSM_T)</td>
<td>SC5</td>
</tr>
<tr>
<td></td>
<td>Are selected because they provide opportunities for students to explain the mathematics behind an answer (FSM_A)</td>
<td>SC6</td>
</tr>
<tr>
<td></td>
<td>Encourage discussion of the connections between various models and strategies (PSC_T)</td>
<td>SC7</td>
</tr>
<tr>
<td></td>
<td>Analyze the connections between various models and procedures (PSC_S)</td>
<td>SC8</td>
</tr>
<tr>
<td>Transmission items</td>
<td>Demonstrate for the class the correct way to use a particular procedure or model before they start solving problems (TSIS_T)</td>
<td>T1</td>
</tr>
<tr>
<td></td>
<td>Present one standard method of solving a task or performing an algorithm (EMRS_T)</td>
<td>T2</td>
</tr>
<tr>
<td></td>
<td>Explain the steps to a procedure or algorithm when I introduce new topics (FSM_T)</td>
<td>T3</td>
</tr>
<tr>
<td></td>
<td>Take notes on how to perform each step in a procedure or algorithm (FSM_S)</td>
<td>T4</td>
</tr>
<tr>
<td></td>
<td>Learn by copying down examples from a teacher demonstration (PSC_S)</td>
<td>T5</td>
</tr>
<tr>
<td></td>
<td>Avoid student errors and misconceptions when a topic is first introduced by explaining how to solve a problem before they start (AM_T)</td>
<td>T6</td>
</tr>
</tbody>
</table>
and transmission-based practice scale scores) and scores based on observations of their practice?

**Data Source**

The data for study two consisted of a sample of 39 fourth-through eighth-grade teachers, who were in their 3rd year of a professional development grant around the DMT professional development framework (Brendefur, Thiede, Strother, Bunning, & Peck, 2013). The MTI course, previously described, was an outgrowth of this much more intensive and sustained professional development project, which involved summer coursework, planning meetings, and in-class support over 3 years. As part of the professional development, teachers were observed once during the fall of the 3rd year of the project, and they completed the instructional practice survey in the winter of the same year. Teachers’ classroom practice was evaluated using a DMT observation instrument built around the DMT framework (Carney, Brendefur, & Hughes, 2014). Similar to our findings with existing instructional practice surveys, existing observation measures (e.g., Mathematical Quality of Instruction, Reformed Teaching Observation Protocol, and Instructional Quality Assessment) framed classroom instructional practice in similar ways but were different enough that they did not capture the full breadth of change we envisioned as a result of our professional development work.

The five domains of the DMT framework provided the structure for the observation instrument development. Each DMT domain was measured with four items. Each item had an overall descriptor followed by specific item descriptors for each score, which ranged from 1–5. The four item scores across each of the five domains were compiled into an overall average for instructional practice on a scale of 1–5, representing the teachers’ level of engagement in the elements of the DMT framework (1 = unaffected, 2 = developing, 3 = engaged, 4 = accomplished, 5 = reflective). The small number of items per domain did not support individual analyses of each domain. The DMT instrument has been shown to have high internal reliability (α = .89 to .98).

**Data Analysis**

Pearson’s *r* was calculated to examine the relationship between the constructs in the instructional practice survey and the teachers’ scores from the DMT classroom observation instrument. High scores on the observation instrument indicate teachers engaging in the DMT instructional practices at higher levels of quality and quantity, whereas high scores on the survey instrument indicate teachers’ reporting high frequency of engagement in the DMT instrument practices.

**Results**

The correlation analysis indicated moderate to high levels of correlation between the DMT observation instrument and the two constructs from the instructional practice survey, social-constructivist and transmission-based practice (see Table 5). The social-constructivist variable positively correlates (*r* = .37, *n* = 39, *p* < .05) and the transmission-based variable negatively correlates (*r* = −.45, *n* = 39, *p* < .05) with teachers’ score on the DMT instructional practice observation instrument. This indicates a relationship between teachers’ self-reports of frequency of instructional practice with observations of their practice conducted by others. In other words, this provides

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Social-constructivist</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Transmission-based</td>
<td>-0.26</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3. DMT observation score</td>
<td>.37*</td>
<td>-.45**</td>
<td>1</td>
</tr>
<tr>
<td><em>M</em></td>
<td>5.97</td>
<td>3.74</td>
<td>56.92</td>
</tr>
<tr>
<td><em>SD</em></td>
<td>0.66</td>
<td>1.53</td>
<td>18.06</td>
</tr>
<tr>
<td>Range</td>
<td>1–7</td>
<td>1–7</td>
<td>20–100</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).
initial evidence that the survey scale scores reflect teachers’ relative frequencies of these instructional practices. Scatterplots of these data are provided in Appendix G.

Discussion and Implications

The initial psychometric evaluation of the survey items revealed several findings. First, the constructs of social-constructivist and transmission-based instructional practices were supported by both the exploratory and confirmatory factor analyses providing evidence of internal structure. In addition, the lack of a moderate to large negative correlation relationship between the social-constructivist and the transmission-based constructs is important to consider given the opposing ways these variables are often presented in the literature and treated in survey measures. The analysis of the retrospective change in instructional practice from prior to after course participation provides initial evidence that the survey scales capture change as a result of particular professional development activities but also reveal an issue with a ceiling effect in the frequency scale. Last, Study 2 provides initial concurrent validity evidence of the relationship between teachers’ self-reported frequency and observations of instructional practice. The following section discusses each of these specific findings followed by more general implications and conclusions.

The two constructs of social-constructivist and transmission-based practice were supported by the exploratory and confirmatory factor analyses. The social-constructivist based practice scale was made up of eight items with two items each coming from four of the five domains from the DMT framework: four from a teacher perspective, two from a student perspective, and two from a tasks and activities perspective. Based on the extent to which the final survey items reflected the original framework, we have reasonable content validity evidence that the final scale reflects the depth and breadth of the social-constructivist based aspects of the DMT framework. In other words, our original theoretical framework from Table 2 was reasonably represented by the eight items remaining following the EFA. However, the lack of items representing the addressing misconceptions domain is concerning. This could be interpreted as a result of a lack of consistent focus within our professional development activities in clearly articulating these ideas. However, there may be other potential causes for the lack of cohesiveness with these items and the social-constructivist based practice scale as operationalized (e.g., participants’ readiness to benefit). Future scale development work will focus on investigating the addressing misconceptions domain and its relationship to the other constructs.

The transmission-based scale consists of six items, one item from each of the DMT framework domains except for focusing on the structure of the mathematics, which has two items. The majority of the items came from a teacher-based perspective (four items), with the other two items coming from the student-based perspective. Again, this provides reasonable content validity evidence that the survey items appear to clearly capture a transmission-based perspective of mathematics instruction. In other words, the six items representing transmission-based practice reasonably represent the breadth of the theoretical framework. However, further scale development will focus on determining whether there are potential tasks or activities that relate specifically to this construct.

The lack of a moderate to strong negative correlation between the transmission and social-constructivist items is a particularly interesting finding. These constructs are often treated as opposing ideas within the mathematics education literature and in particular in measures of teachers’ instructional practice. For example, Swan (2006) reverse scored the student-centered items and placed them on the same scale as the teacher-centered items. Ross and colleagues (2003) used a similar design. This assumes unidimensionality and a polar relationship, which we found not to be the case. In other words, it appears from our data that social-constructivist and transmission-based instructional practice are not opposite ends of a continuum and should be examined as two separate variables. Whether our results are particular to our sample or whether this is a consistent finding across teachers requires further investigation but should make researchers consider analyzing the structure of data from items that are typically reverse scored.

The evidence related to the ability of the items to measure change in instructional practice as a result of professional development shows promise. Our findings indicate a significant change in the participants’ perceptions of their instructional practice from prior to after using a retrospective model. Further investigation needs to occur to determine how closely survey responses match observed changes in practice and whether the internal structure holds beyond the bounds of our particular professional development project. In addition, the ceiling effect in the frequency scale may indicate a need to expand the current scale in future studies.

Study 2 examined the concurrent validity between the self-reported survey constructs and observations of instructional practice. The moderately to strongly significant correlations in the expected directions provide promising validity evidence regarding the accuracy of teachers’ self-reported frequency of instructional practice. In other
words, it provides evidence that teachers’ self-reported instructional practices are moderately correlated with their actual practices. The DMT observation instrument is designed to capture the quality of mathematics instructional practice, while our survey instrument is designed to capture the frequency of those practices. The fact that we were able to capture moderate to strong relationships between these variables with related but different scale foci is promising. The evidence of this relationship should continue to be examined in future studies.

Lastly, what does all this mean to mathematics teacher educators looking for a survey scale to measure instructional practice? The validity evidence from each survey scale must be examined in relation to the context in which it will be used, such as the specific goals of the professional development program, how the constructs within a survey are defined and operationalized, and the complexities of creating and validating a survey scale. As a research community, mathematics teacher educators need a range of scales to select from and need to understand the various pros and cons of each. The following ideas are provided to stimulate thinking about various considerations when selecting a survey scale.

The TIMSS-Engage and Horizon-Reform survey scales are particularly useful for researchers who want to compare their results with an international or national sample. However, researchers using these scales need to closely examine how the scales conceptualize instructional practice to determine whether that conceptualization matches the research project’s needs. The Swan-Practices and Ross-Standards scales provide a much larger item bank in their survey scales and therefore may provide a broader picture of instructional practice. However, researchers who use these scales may want to utilize factor analysis to examine the unidimensionality of these scales. These scales offer more detailed evidence of the relationship of their scale scores to other variables of potential interest. However, this information needs to be carefully considered in terms of its generalizability to other situations.

Our instructional practice instrument provides separate scales for social-constructivist and transmission-based variables. This allows researchers to evaluate the level of each variable separately within their sample. In addition, initial evidence supports the use of this instrument on a large scale to investigate change in instructional practice as a result of professional development focused on aspects of the DMT framework.

Through the lens of our own professional development project, our current results indicate we increased the frequency with which participants engaged in social-constructivist practices and decreased the frequency with which participants engaged in transmission-based practices. The results of Study 2 provide initial support for the claim that these self-reported changes in practice translated into actual classroom practice. Given the large-scale nature of our project—to date over 12,000 participants have completed the course—this provides promising evidence that high-quality, large-scale mandated professional development has the potential to shift classroom practice.

Additional validation work needs to be conducted to determine the usefulness of the instrument across varied settings (e.g., outside of Idaho) and participants (e.g., preservice teachers) and the relationship between the survey scale scores to other variables and/or constructs of importance (e.g., student achievement). For example, examining the relationship between teachers’ reported frequency of instructional practice for each construct in relation to measures of student achievement or socioemotional well-being has the potential to provide quantitative evidence for the frequency with which teachers should engage in different types of instructional practice. The current trend appears to support increasing social-constructivist practices and decreasing transmission-based practices, but perhaps how these two modes interplay in the classroom might give us more useful information. In addition, our focus on frequency versus quality of instructional practice should be kept in mind. How might this impact findings? We welcome careful use, modification, and further study of this instrument and hope it serves to spark further discussion around ideas of measuring practice on a large scale.

References


**Authors**

Michele B. Carney, PhD, 1910 University Drive, Boise State University, Boise, ID 83725-1745; michelecarney@boisestate.edu

Jonathan L. Brendefur, PhD, 1910 University Drive, Boise State University, Boise, ID 83725-1745; jonathanbrendefur@boisestate.edu

Gwyneth R. Hughes, 1910 University Drive, Boise State University, Boise, ID 83725-1745; gwynethhughes@boisestate.edu

Keith Thiede, 1910 University Drive, Boise State University, Boise, ID 83725-1745; keiththiede@boisestate.edu
Appendix A: Mathematical Instructional Practice

### Mathematics Instructional Practice - Tasks and Activities

Indicate for each statement the frequency you engage in the particular instructional practice both *prior* to and *after* participation in the MTI course.

#### Classroom tasks and activities:

<table>
<thead>
<tr>
<th>Activity</th>
<th>BEFORE Participation in MTI</th>
<th>AFTER Participation in MTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are selected because they provide opportunities for students to explain the mathematics behind an answer.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are directed by the sequence of the textbook</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are based on the connections that can be made between various models and algorithms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Involve the intentional presentation of solution strategies containing misconceptions and/or mistakes for student to diagnose and correct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are selected because the problem’s context (or situation) elicits a particular model or models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are selected because they allow students repeated practice to learn a procedure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus on repeated practice of a model or procedure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus on mathematical procedures in order to avoid confusion and prevent student errors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primarily focus on learning a particular procedure or algorithm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are based on their potential to encourage discussions of students’ mathematical ideas</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A drop-down box for each question lists the following options:

- Never
- 2–3 times per year
- Once per month
- 2–3 times per month
- Once a week
- 2–3 times per week
- Daily
Appendix B: Exploratory Factor Analysis

Exploratory factor analysis. For each anticipated latent variable, an EFA using principal components extraction method with varimax rotation was conducted related to instruction both prior to and after the MTI course (using SPSSv21). The first EFA analyzed (1) 15 social-constructivist items measuring prior practice and (2) the same 15 social-constructivist items measuring after practice, and the second EFA analyzed (3) 15 transmission-based items measuring practice prior and (4) the same 15 items measuring transmission-based practice after.

The factor loadings and summary statistics are presented for the social-constructivist (Table B1) and transmission (Table B2) items. For social-constructivist, two factors were extracted with eigenvalues over one, 6.595 and 1.211 for prior, and three factors with eigenvalues over one, 5.895, 1.503, and 1.129 for after. Eight items had factor loadings above .40, loaded on the same component both prior and after, and had no cross-loading. These eight items were identified as measuring the latent variable social-constructivist practice and retained for analysis in the CFA.

For transmission, two factors were extracted with eigenvalues over one, 6.808 and 1.459 for prior, and three factors with eigenvalues over one, 6.946, 1.271, and 1.003 for after. Component one had six items with factor loadings over .40 and loaded only on component one both prior and after with no cross-loading. These six items were identified as measuring the latent variable of transmission-based instructional practice and retained for analysis in the CFA. Due to the dramatic drop in eigenvalues after component one, the remaining items were dropped from further analysis (Costello & Osborne, 2005).

### Table B1
Rotated Component Matrix for Social-Constructivist Items

<table>
<thead>
<tr>
<th>DMT framework element</th>
<th>Prior components</th>
<th></th>
<th></th>
<th>After components</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>TSIS_T</td>
<td>.712</td>
<td>-.063</td>
<td>.734</td>
<td>.033</td>
<td>.252</td>
<td></td>
</tr>
<tr>
<td>TSIS_A</td>
<td>.736</td>
<td>.190</td>
<td>.750</td>
<td>.068</td>
<td>.172</td>
<td></td>
</tr>
<tr>
<td>EMRS_T</td>
<td>.768</td>
<td>.026</td>
<td>.731</td>
<td>.066</td>
<td>.027</td>
<td></td>
</tr>
<tr>
<td>EMRS_S</td>
<td>.797</td>
<td>.072</td>
<td>.705</td>
<td>.158</td>
<td>.225</td>
<td></td>
</tr>
<tr>
<td>FSM_T</td>
<td>.750</td>
<td>-.055</td>
<td>.750</td>
<td>.159</td>
<td>.162</td>
<td></td>
</tr>
<tr>
<td>FSM_A</td>
<td>.748</td>
<td>.239</td>
<td>.780</td>
<td>.081</td>
<td>.019</td>
<td></td>
</tr>
<tr>
<td>PSC_T</td>
<td>.748</td>
<td>.067</td>
<td>.631</td>
<td>.117</td>
<td>.217</td>
<td></td>
</tr>
<tr>
<td>PSC_S</td>
<td>.801</td>
<td>.190</td>
<td>.706</td>
<td>.257</td>
<td>.168</td>
<td></td>
</tr>
<tr>
<td>EMRS_A</td>
<td>.159</td>
<td>.757</td>
<td>.161</td>
<td>.632</td>
<td>-.265</td>
<td></td>
</tr>
<tr>
<td>FSM_S</td>
<td>.576</td>
<td>.174</td>
<td>.212</td>
<td>.622</td>
<td>.249</td>
<td></td>
</tr>
<tr>
<td>PSC_A</td>
<td>.575</td>
<td>.381</td>
<td>.342</td>
<td>.682</td>
<td>.120</td>
<td></td>
</tr>
<tr>
<td>AM_T</td>
<td>.520</td>
<td>.002</td>
<td>.212</td>
<td>-.024</td>
<td>.783</td>
<td></td>
</tr>
<tr>
<td>AM_A</td>
<td>.573</td>
<td>.266</td>
<td>.252</td>
<td>.201</td>
<td>.735</td>
<td></td>
</tr>
<tr>
<td>AM_S</td>
<td>.670</td>
<td>.122</td>
<td>.574</td>
<td>.168</td>
<td>.441</td>
<td></td>
</tr>
<tr>
<td>Eigenvalues</td>
<td>6.595</td>
<td>1.211</td>
<td>5.895</td>
<td>1.503</td>
<td>1.129</td>
<td></td>
</tr>
<tr>
<td>Variance explained</td>
<td>44.0%</td>
<td>8.1%</td>
<td>39.3%</td>
<td>10.0%</td>
<td>7.5%</td>
<td></td>
</tr>
</tbody>
</table>
Table B2
Rotated Component Matrix for Transmission Items

<table>
<thead>
<tr>
<th>DMT framework element</th>
<th>Prior components</th>
<th>After components</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>TSIS_T</td>
<td>.679</td>
<td>.378</td>
</tr>
<tr>
<td>EMRS_T</td>
<td>.731</td>
<td>.249</td>
</tr>
<tr>
<td>FSM_T</td>
<td>.717</td>
<td>.355</td>
</tr>
<tr>
<td>FSM_S</td>
<td>.528</td>
<td>.120</td>
</tr>
<tr>
<td>PSC_S</td>
<td>.701</td>
<td>.203</td>
</tr>
<tr>
<td>AM_T</td>
<td>.716</td>
<td>.296</td>
</tr>
<tr>
<td>EMRS_S</td>
<td>.213</td>
<td>.795</td>
</tr>
<tr>
<td>EMRS_A</td>
<td>.276</td>
<td>.743</td>
</tr>
<tr>
<td>PSC_A</td>
<td>.222</td>
<td>.763</td>
</tr>
<tr>
<td>TSIS_A</td>
<td>.441</td>
<td>.269</td>
</tr>
<tr>
<td>PSC_T</td>
<td>.543</td>
<td>.422</td>
</tr>
<tr>
<td>FSM_A</td>
<td>.692</td>
<td>.431</td>
</tr>
<tr>
<td>AM_A</td>
<td>.609</td>
<td>.483</td>
</tr>
<tr>
<td>AM_S</td>
<td>.331</td>
<td>.696</td>
</tr>
<tr>
<td>TSIS_S</td>
<td>-.553</td>
<td>.390</td>
</tr>
</tbody>
</table>

Eigenvalues
- 6.808
- 1.459
- 6.946
- 1.271
- 1.003

Variance explained
- 45.4%
- 9.7%
- 46.3%
- 8.5%
- 6.7%

TSIS = Taking students’ ideas seriously; PSC = Pressing students conceptually; FSM = Focusing on the structure of mathematics; AM = Addressing misconceptions; EMRS = Encouraging multiple representations and strategies
A = Activities (and tasks); T = Teacher; S = Student

(Return to page 101)
## Appendix C: Mathematics Instructional Practice Survey

### Mathematics Instructional Practice Survey

Indicate for each statement the frequency you engage in the particular instructional practice.

### As the classroom teacher, I:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Daily</th>
<th>2-3 times per week</th>
<th>Once per week</th>
<th>Once per month</th>
<th>Once per year</th>
<th>Never</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emphasize the use of multiple models for recording and communicating student thinking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Encourage discussion of the connections between various models and strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilitate discussion about underlying mathematical concepts (e.g., composing or decomposing number)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present one standard method of solving a task or performing an algorithm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilitate small group or whole class discussion on student thinking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explain the steps to a procedure or algorithm when I introduce new topics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demonstrate for the class the correct way to use a particular procedure or model before they start solving problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avoid student errors and misconceptions when a topic is first introduced by explaining how to solve a problem before they start</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Students:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Daily</th>
<th>2-3 times per week</th>
<th>Once per week</th>
<th>Once per month</th>
<th>Once per year</th>
<th>Never</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyze the connections between between various models and procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Take notes on how to perform each step in a procedure or algorithm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learn by copying down examples from a teacher demonstration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solve problems that allow for several different approaches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Classroom tasks and activities:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Daily</th>
<th>2-3 times per week</th>
<th>Once per week</th>
<th>Once per month</th>
<th>Once per year</th>
<th>Never</th>
<th>Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are based on their potential to encourage discussions of students’ mathematical ideas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are selected because they provide opportunities for students to explain the mathematics behind an answer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table D1
Item Correlations, Means, and Standard Deviations for Prior to MTI Course Participation

<table>
<thead>
<tr>
<th>Correlations</th>
<th>Social-constructivist items</th>
<th>Behaviorist</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SC1</td>
<td>SC2</td>
</tr>
<tr>
<td>SC1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC2</td>
<td>.524**</td>
<td></td>
</tr>
<tr>
<td>SC3</td>
<td>.512**</td>
<td>.535**</td>
</tr>
<tr>
<td>SC4</td>
<td>.503**</td>
<td>.568**</td>
</tr>
<tr>
<td>SC5</td>
<td>.501**</td>
<td>.446**</td>
</tr>
<tr>
<td>SC6</td>
<td>.482**</td>
<td>.677**</td>
</tr>
<tr>
<td>SC7</td>
<td>.515**</td>
<td>.491**</td>
</tr>
<tr>
<td>SC8</td>
<td>.492**</td>
<td>.554**</td>
</tr>
<tr>
<td>T1</td>
<td>-.146**</td>
<td>-.171**</td>
</tr>
<tr>
<td>T2</td>
<td>-.252**</td>
<td>-.297**</td>
</tr>
<tr>
<td>T3</td>
<td>-.142**</td>
<td>-.182**</td>
</tr>
<tr>
<td>T4</td>
<td>-.03</td>
<td>-.08</td>
</tr>
<tr>
<td>T5</td>
<td>-.243**</td>
<td>-.272**</td>
</tr>
<tr>
<td>T6</td>
<td>-.229**</td>
<td>-.220**</td>
</tr>
</tbody>
</table>

Mean: 4.75  4.25  4.20  4.51  3.97  4.24  4.37  4.09  5.86  5.09  5.69  4.62  5.31  5.27
SD: 1.90  2.08  2.06  1.95  2.16  2.04  2.06  2.05  1.60  2.08  1.72  2.25  2.04  2.06
Table D2
Item Correlations, Means, and Standard Deviations for After MTI Course Participation

<table>
<thead>
<tr>
<th>Correlations AFTER</th>
<th>Social-constructivist items</th>
<th>Behaviorist</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SC1</td>
<td>SC2</td>
</tr>
<tr>
<td>SC1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>SC2</td>
<td>.508**</td>
<td>1</td>
</tr>
<tr>
<td>SC3</td>
<td>.493**</td>
<td>.517**</td>
</tr>
<tr>
<td>SC4</td>
<td>.507**</td>
<td>.594**</td>
</tr>
<tr>
<td>SC5</td>
<td>.551**</td>
<td>.569**</td>
</tr>
<tr>
<td>SC6</td>
<td>.530**</td>
<td>.689**</td>
</tr>
<tr>
<td>SC7</td>
<td>.511**</td>
<td>.472**</td>
</tr>
<tr>
<td>SC8</td>
<td>.487**</td>
<td>.589**</td>
</tr>
<tr>
<td>T1</td>
<td>-.118*</td>
<td>-.138**</td>
</tr>
<tr>
<td>T2</td>
<td>-.102*</td>
<td>-.09</td>
</tr>
<tr>
<td>T3</td>
<td>-.09</td>
<td>-.09</td>
</tr>
<tr>
<td>T4</td>
<td>-.04</td>
<td>-.03</td>
</tr>
<tr>
<td>T5</td>
<td>-.159</td>
<td>-.195</td>
</tr>
<tr>
<td>T6</td>
<td>-.134**</td>
<td>-.08</td>
</tr>
<tr>
<td>SD</td>
<td>0.90</td>
<td>1.18</td>
</tr>
</tbody>
</table>
Appendix E: Confirmatory Factor Analysis

Confirmatory factor analysis. The purpose of confirmatory factor analysis is to determine the goodness of fit between a hypothesized model and the sample data. Our hypothesized a priori structure is comprised of (a) two factors, social-constructivist (soccon) and transmission (trans), (b) with each item having a nonzero loading on the factor it is designed to measure and a zero loading on factors it is not designed to measure, and (c) the residuals for each item are uncorrelated with each other. The next stage of analysis involved using the second half of the randomly split data to examine the fit of this model \( (n = 399) \). The MLM estimator in Mplus7 was utilized due to issue with multivariate non-normality, requiring listwise deletion of missing data and reducing the sample size \( (n = 326 \text{ prior}, n = 319 \text{ after}) \). Our initial analysis indicated a potential model misspecification, and our model was adjusted post hoc to allow for error covariance between items T2 and T6, T4 and T5, and SC2 and SC6. Due to the conceptual similarity in the items, this adjustment still fit our initial framework. Figure E1 contains the full diagram results of the final CFA, and Table 5 provides the fit statistics. While there is no clear consensus regarding the indices that are most appropriate for model fit, our indices fit all the major recommendations (Byrne, 2013).

**Table 5**
Fit Statistics for Final CFA Model with MLM Estimator for Prior and After MTI Course Participation

<table>
<thead>
<tr>
<th>Model</th>
<th>( N )</th>
<th>( \chi^2 )</th>
<th>( df )</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>326</td>
<td>131.30*</td>
<td>73</td>
<td>.049</td>
<td>.067</td>
<td>.970</td>
</tr>
<tr>
<td>After</td>
<td>319</td>
<td>101.56*</td>
<td>73</td>
<td>.035</td>
<td>.044</td>
<td>.972</td>
</tr>
</tbody>
</table>

* \( p < .01 \)

Two key findings emerged from the CFA. First, the factor structure resulting from the EFA was supported by the CFA model. In other words, the survey item responses loaded on the latent factors as expected. Second, the construct of social-constructivist instructional practice was found to be only negligibly correlated to transmission-based instructional practice for prior \( r = -.18 \) and not significantly correlated for after \( r = -.16 \). The lack of a moderate to strong negative correlation between the two latent variables indicates our survey participants view these as distinct and unrelated constructs within mathematics instructional practice.
Figure E1. CFA for responses related to instruction *prior* to the MTI course.
Figure E2. CFA for responses related to instruction after the MTI course.
Appendix F: Histograms of Scale Scores

**Figure F1.** Histograms of social-constructivist scale scores for prior and after MTI course participation.

**Figure F2.** Histograms of transmission-based scale scores for prior and after MTI course participation.
Appendix G: Correlation of Scale Scores to DMT Observation

Figure G1. Scatterplot of relationship between social-constructivist and the DMT observation instrument scores $r = (37) .37, p < .05$.

Figure G2. Scatterplot of relationship between transmission and the DMT observation instrument scores $r = (37) -.45, p < .01$.

1 = Never
2 = 2–3 times per year
3 = Once per month
4 = 2–3 times per month
5 = Once a week
6 = 2–3 times per week
7 = Daily
CAEP Program Reviewer Recognition

The National Council of Teachers of Mathematics expresses its appreciation to the following for their roles as program reviewers in the 2014–2015 academic year for the Council for the Accreditation of Educator Preparation:

Linda Arnold
Julie Belock
Elizabeth M. Brown
Suzanne Brown
Christine A. Browning
Joanne C. Caniglia
Gregory Chamblee
Ted Coe
Thomas E. Cooper
Beth Greene Costner
Kelly M. Costner
Irma Cruz-White
Lisa Darlington
Richard DeCesarer
Dana Dodson
Gina Dunn
Nancy Tanner Edwards
Thomas A. Evitts
Francis M. Fennell
Carol Fowlkes
Joanne E. Goodell
Marilyn Hala
Shannon Harmon
Michele A. Iiams
Pat Lamphere Jordan
Karen D. King
Marshall B. Lassak
Michael Long
Monique C. Lynch
Kathleen Lynch-Davis
Alison Mall
Joanna O. Masingila
Edna Meisel
Roxanne V. Molina
Judy O’Neal
Karla J. Oty
Sarah E. Patrick
Alonzo F. Peterson
Jana Peterson
Marnie C. Phipps
Terrie T. Poehl
Andrew Polly
Ernest O. Pratt
David K. Pugalee
Christopher Rakes
Katharine D. Rasch
Ann Calder Rose
Connie Schrock
Kathryn G. Shafer
Jeanne Shimizu
Betsy Showalter
Kelli Slaten
Beverly Smith
Ron Smith
L. Lynn Stallings
Michael D. Steele
Paula R. Stickles
Sylvia Taube
James Telese
Sandra Trotman
Carol Wade
Patricia Wagner
Crystal Y. Walcott
Thomas P. Walsh
Judy A. Werner
Trena L. Wilkerson
Kay Wohlhuter