Volume 3 • Number 2 • March 2015

Mathematics Teacher Educator

EDITORIAL

ARTICLES

National Council of Teachers of Mathematics • Association of Mathematics Teacher Educators

Mathematics Teacher Educator

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.

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 NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS

AMTE Association of Mathematics Teacher Educators

EDITORIAL

Reflections on the First 3 Years of *Mathematics Teacher Educator*: Successes and Challenges

Margaret S. Smith
Editor, Mathematics Teacher Educator

As I indicated in my editorial in the inaugural issue of *MTE*, the publication of *MTE* represented an historical moment for mathematics teacher education, providing for the first time a much-needed forum for supporting and improving the practice of educating teachers of mathematics. This current issue of *MTE*, the last one for which I will serve as editor, seems an appropriate time to reflect on the successes of the journal to date and the challenges that the editorial board is actively addressing.

Successes

Most notably, the 29 articles that have been published over the past 3 years (including those found in the current issue) provide the starting point for the knowledge base for mathematics teacher education that was conceived by the National Council of Teachers of Mathematics (NCTM) and Association of Mathematics Teacher Educators (AMTE) visionaries who laid the groundwork for the journal. As the Editorial Board stated in the call for manuscripts, first published in 2011, "The journal is a tool to build the personal knowledge that mathematics educators gain from their practice into a trustworthy knowledge base that can be shared with the profession." While we have yet to see articles that explicitly build on work that has previously been published in the journal, the foundation for such building has been laid with articles that address a breadth and depth of issues faced by mathematics teacher educators in their work.

Many of the articles that have been published in *MTE* have taken advantage of the online nature of the journal. For example, in the article "Using Simulations to Foster Preservice Mathematics Teachers' Self-Assessment, Learning, and Reflections on Teaching" (Volume 1, Issue 2), Garofalo and Trinter (2013) link to a movie of one of the simulations created by a preservice teacher. In the article "Developing Teachers' Knowledge of a Transformations-Based Approach to Geometric Similarity" (Volume 2, Issue 1), Seago and her colleagues (2013) include dynamic figures that make the relationships they describe salient. In addition, they link to video clips of students applying transformation-based and static perspectives in solving a geometry task. In both these cases, the technology serves to enhance the article and make the content more accessible to the reader.

Other authors have taken advantage of the journal's online presence by providing additional content. For example, in the article "Using 'Lack of Fidelity' to Improve Teaching" (Volume 1, Issue 1), Morris (2012) includes an example of the type of detailed lesson plan she refers to in the article. Not only does the inclusion of the lesson plan allow the reader to see exactly what she is describing, but the link within the article makes it possible to view the lesson plan when it is first mentioned and then return to the main article. In addition, teacher educators might choose to use or adapt the lesson plan in their own work. While we may have only scratched the surface in terms of what is possible with our online format, these examples highlight some of the possibilities.

A key element in building the desired knowledge base is helping educate authors and reviewers about the nature of the journal and the characteristics of articles that are consistent with our mission. We have operationalized this educative function primarily by providing feedback to authors and reviewers. All authors are given detailed feedback on how to improve their manuscripts, regardless of whether the author received a decision of accept, reject, or revise and resubmit. For manuscripts that have been rejected, reviewers make an effort to clearly explain the decision to the author and to cite specific ways in which the manuscript could be improved. I have been pleasantly surprised at the number of authors who have written to express their appreciation for this feedback even though their manuscripts were not accepted. Below are two examples from emails I have received (reprinted with permission from the authors).

Thank you for the feedback regarding my submitted article. The feedback was very helpful and I think it has given me a better sense of the scope of the journal as a whole. I do not know if you typically do things like this, but I would appreciate you sending my particular thanks along to reviewer 2. His/Her comments were very insightful and helpful in moving forward. I plan on composing another article framing this work from a different perspective and providing more emphasis on the details of my study and the evidence. I hope that it will be a better fit for the journal. [*Author 1*]

Thank you very much for this notification. I appreciate your time, the reviewers' time, and the thoughtful response you composed and reviews you forwarded. I

look forward to reworking the paper given this helpful feedback. [*Author 2*]

Every manuscript is reviewed by three people—two selected from the reviewer database who have some level of expertise with the topic addressed in the manuscript, and one member of the editorial panel. All reviewers receive copies of the decision letter that is sent to the author as well as copies of all three of the reviews. The purpose of this is to help reviewers improve their reviews. Although this feedback is less direct than what is provided to authors, notes from two different reviewers provide evidence that this process has helped reviewers reflect on the quality of their reviews. One reviewer (who wrote a very short review with very little detail) wrote to say that after reading the other reviews and the decision letter for the manuscript she had reviewed, she had a better understanding of what was expected and hoped she would be given another chance to review. In a second case, after receiving and reading the other reviews and the decision letter, another reviewer wrote to say that she realized that she needed to "step up her game."

In addition to receiving reviews and decision letters for manuscripts they have reviewed, members of the editorial panel engage in a group review of one or two manuscripts each year in order to ensure consistency in ratings across members of the panel. This calibration activity provides an opportunity for members of the panel to discuss the review criteria and how they are applied to particular manuscripts. New members of the board receive additional training through a webinar conducted each spring that orients them to the roles and responsibilities of members of the editorial panel.

Challenges

The key challenge the journal is facing is the relatively small number of submitted manuscripts (30 out of 294) that have met the established standards for the journal. While the journal can accommodate as many as six manuscripts in a single issue, this has only been accomplished twice (Volume 1, Issue 2; Volume 2, Issue 1), and we have published as few as three manuscripts in an issue (Volume 3, Issue 1). The editorial board has taken several steps to address the issue of manuscript flow and quality: 1) make criteria for the journal more transparent; 2) feature successful authors at conference sessions who can "tell their stories"; 3) put out a call for manuscripts on a particular topic; and 4) invite manuscripts.

Make criteria more transparent. The Editorial Board is working to make the criteria for manuscripts more transparent through two efforts. First, the call for manuscripts is being revised to make clearer the types of manuscripts we are looking for and the essential elements of the manuscripts. In particular, we are articulating criteria for manuscripts that do not fall into the category of describing a successful intervention. Second, the editorial feature of the journal has been and will be used to specifically address limitations reviewers frequently identify in submitted manuscripts. Most notably, the editorial "Linking Claims and Evidence" (Volume 1, Issue 2) attempted to articulate the nature of evidence appropriate for *MTE* and provide specific examples of what does and does not constitute sufficient evidence to support a claim. We targeted this particular issue because the most common reason for reviewers to reject a manuscript is a lack of evidence for the authors' claims.

Feature successful authors. The Editorial Board invited Eva Thanheiser to share her and her coauthors' experience with the revise and resubmit process at a session held at the annual AMTE meeting in January 2013. The purpose of this session was to provide a specific example, from the first author's perspective, of the relationship between an original article that is submitted to the journal, the feedback that is sent to the authors, and the final version of the article. The intent was to encourage authors not to give up when asked to revise a manuscript and to seriously consider the feedback that is provided. (See the editorial in Volume 2, Issue 1 for a description of this session and Volume 1, Issue 2 for the article.) Given the positive reaction to this session, the Editorial Board decided to showcase additional authors in 2015. At the annual AMTE meeting in February 2015, Michael Steele and Amy Hillen discussed how they turned a 2011 AMTE presentation into a published article (which appeared in Volume 1, Issue 1 of *MTE*). Similarly, at the NCTM Research Conference in April 2015, Kevin Moore will describe how he and his colleagues turned their research work into a successful *MTE* submission that addressed a teacher education audience (see Volume 2, Issue 2).

Call for manuscripts on a specific topic. Calls for manuscripts that focus on particular topics and have specific deadlines can provide a focus for potential authors. Toward this end, the Board developed a call for manuscripts related to *Principles to Actions: Ensuring Mathematical Success for All* (NCTM, 2104). This call was introduced in the summer 2014 issue of *Connections*, appeared in the September 2014 issue of *MTE*, and was posted on the AMTE and NCTM websites. The call invites authors to submit manuscripts by September 1, 2015, that address the following questions: How can *Principles to Actions: Ensuring Mathematical Success for All* be used to design learning experiences for teachers? What impact do these experiences have on teachers and their students?

Invite manuscripts. To augment the number of articles submitted to the journal and hence provide more content to readers in each issue of *MTE*, the Board decided to invite authors to submit manuscripts. According to the Handbook for the Editorial Board for *MTE* (p. 15):

The editor is authorized to invite manuscripts for the journal to highlight particular issues in mathematics teacher education or for other reasons that may enhance the journal. The selection of authors and topics is at the discretion of the editor. In general, no more than one invited piece should be published in each issue (although two shorter pieces might be published in a point/counterpoint fashion).

Any manuscript that is invited will be sent to two editorial panel members who will provide feedback to strengthen the manuscript. This review will not be blind and will be done outside of the manuscript system. Invited manuscripts will not be included in journal statistics related to acceptance rate or time to decision.

The first invited manuscript, written by Melissa Boston, Jonathan Bostic, Kristen Lessig, and Milan Sherman, appears in this volume of the journal. The article, "A Comparison of Mathematics Classroom Observation Protocols," is based on a session that the four coauthors presented at the 2014 AMTE meeting. The article was invited for several reasons: 1) the session was very popular, with standing room only, and hence was of interest to AMTE members; 2) the topic, observation tools, continues the focus on tools that was first introduced in a *MTE* editorial (Volume 3, Issue 1); and 3) the author team provides a example of a partnership between more and less experienced authors. In fact, the analysis of classroom observation tools described in this manuscript began within a working group in the Service, Teaching, and Research Program (STaR) in Mathematics Education in which the second, third, and fourth authors participated.

Looking Forward

While launching a new journal is not without challenges, there are lots of reasons to be optimistic about its future. The number of manuscripts reviewed increased by more than 20% in the past year. Individual subscriptions to the journal rose more than 70% from 2013 to 2014, while the number of institutional subscriptions more than doubled in the same time period. The mathematics teacher education community has supported the journal by volunteering to review manuscripts, returning prompt and detailed reviews of manuscripts, and attending journal sessions held at annual meetings.

To remain viable, the journal needs more contributions of high-quality manuscripts. Help us increase awareness of the journal by sharing particular articles with colleagues and students. Please read the revised call for manuscripts and think about what you might submit. Also, share the call with your colleagues and graduate students and encourage them to submit manuscripts to the journal. If you are in a position to do so, consider mentoring a colleague or student by coauthoring a manuscript or actively supporting them in writing a manuscript for the journal. Consider also the ways in which the presentation of your work or your colleagues' work could be enhanced by the online nature of this journal.

The journal will continue to grow and develop over the next few years under the leadership of Sandra Crespo (Michigan State University), who will take over as editor of *MTE* in May 2015. Sandra is an accomplished teacher education scholar who will provide new energy and creativity for the journal. I appreciate having had the opportunity to serve as the founding editor of *MTE* and to work with a talented associate editor and dedicated members of the editorial panel. I look forward to seeing how *MTE* grows and evolves toward the goal of building "a trustworthy knowledge base that can be accessed and shared widely in the profession" (Hiebert, Gallimore, & Stigler, 2002, p. 3).

Reference

Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What it would look like and how we can get one. *Educational Researcher*, *31*(5), 3–15.

Author

Margaret S. Smith, Department of Instruction and Learning, University of Pittsburgh, Pittsburgh, PA 15260; pegs@pitt.edu

Multiplication by 10_{five}: Making Sense of Place Value Structure Through an Alternate Base

Jodi Fasteen
Portland State University, Oregon

Kathleen Melhuish
Portland State University, Oregon

Eva Thanheiser
Portland State University, Oregon

Prior research has shown that preservice teachers (PSTs) are able to demonstrate procedural fluency with whole number rules and operations, but struggle to explain why these procedures work. Alternate bases provide a context for building conceptual understanding for overly routine rules. In this study, we analyze how PSTs are able to make sense of multiplication by 10_{five} in base five. PSTs' mathematical activity shifted from a procedurally based concatenated digits approach to an explanation based on the structure of the place value number system.

Keywords: Alternate bases; Place value; Preservice teachers; Whole numbers/Natural numbers/Counting numbers

Mathematical knowledge is based on both convention and logic. However, convention in this case serves as a shelter for those who don't have a conceptual understanding of a mathematical procedure (Ma, 1999, p. 31).

Setting the Stage: An Example for the Reader

When school children encounter base ten numbers, they often have limited knowledge of the structure of our base ten system. To simulate a similar experience for preservice elementary teachers (PSTs), we often ask them to construct numbers in a different base for which they do not have an explicit structure yet in place. To allow the reader to experience some of what the PSTs experience, we begin with an example situated in base five. In base five we collect units until we have 5 units, then we create a group of 5 to form a new unit. We continue this way, grouping 5 of each unit type to form the next larger unit. Consider the number 123_{five}, which refers to 1 group of twenty-five, 2 groups of five and 3 ones (see Figure 1a). To provide some language to talk about the numbers in base five, a group of 5 ones will be referred to as a *long*, a collection of 25 ones (or 5 longs) will be referred to as a *flat,* and a collection of 125 ones (or 5 flats) will be referred to as a *long flat* (see Figure 1b). The numeral 23_{five} can be read aloud as 2 longs 3 ones. While it is possible to convert a base five number like 123_{five} into base ten (it is 38), we ask the reader to take a minute and try to solve the following problem without converting out of base five. This will allow you to experience working in an unfamiliar base before reading the remainder of the paper.

Problem: *Find the product 3_{five} x 23_{five}.*

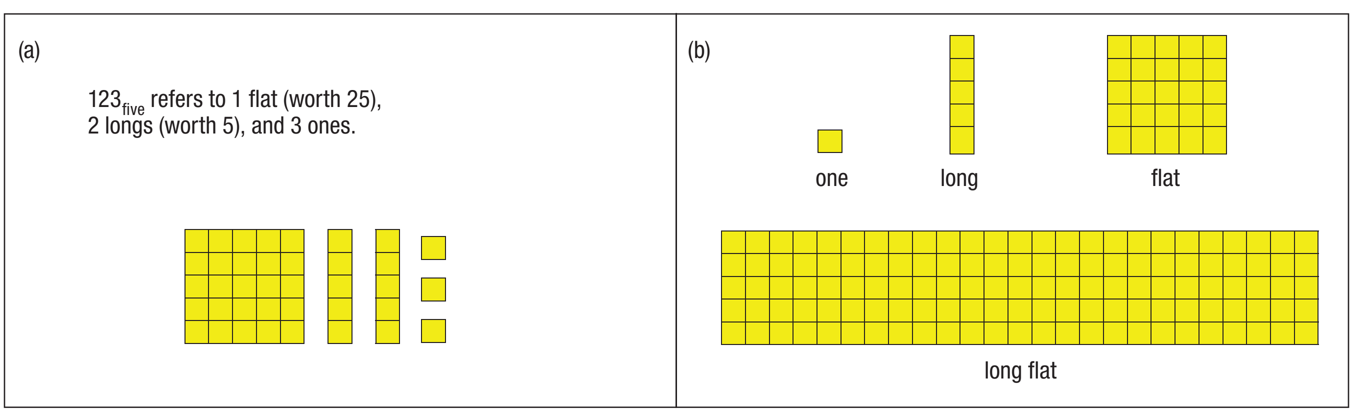


Figure 1. **An introduction to base five language and imagery.**

Introduction

Current educational policy documents in the United States advocate that children develop a conceptual understanding of mathematics, focusing on sense making in addition to procedural fluency (Kilpatrick, Swafford, & Findell, 2001; National Council of Teachers of Mathematics, 2000; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). We will adopt Hiebert and Lefevre's (1986) characterization of conceptual and procedural knowledge. Conceptual knowledge is knowledge where "the linking relationships are as prominent as the pieces of information" (1986, pp. 3–4). For example, conceptual knowledge of base ten might include an understanding of the relationship between adjacent unit types (ones, tens, hundreds), and thus a number like 300 might be understood as being 300 *ones*, 30 copies of *ten*, 3 copies of *hundred,* etc. In contrast, Hiebert and Lefevre identify procedural knowledge as knowledge of symbols, rules, and algorithms. Knowledge that the 3 in 345 is in the hundred's place because that is the name of the place (label) rather than the value would be considered procedural knowledge. In order to support their students' development of conceptual knowledge, teachers also need to have conceptual knowledge of the mathematics they teach.

Preservice teachers (PSTs) often demonstrate the ability to use procedures efficiently but lack the knowledge to explain why these procedures work (Ball, 1988; Lo, Grant, & Flowers, 2008; Ma, 1999; Thanheiser, 2009a). Further, the PSTs may be unaware that there is a rationale (a "why") for each step in the procedure (Thanheiser, Philipp, Fasteen, Strand, & Mills, 2013). However, the PSTs' lack of conceptual knowledge may be masked by their use of mathematical terminology, including procedural use of place value language (Ma, 1999; Thanheiser, 2009a). For example, a PST who is multiplying the multidigit numbers 24 x 38 may state that the product of the 2 and 3 is a 6 that is placed in the hundreds column, using the term *hundreds* as a label, but without understanding that the 2 refers to 2 tens or 20, the 3 refers to 3 tens or 30, and the 6 refers to 6 hundreds.

In this article, we examine how an alternate base context might contribute to the development of conceptual knowledge of the multiplicative structure of place value in a preservice mathematics content course on number and operation. By understanding place value we mean understanding that a digit has a face value (value of a digit) and a place value (value of the unit type the digit represents) and that there is a consistent 10 to 1 multiplicative relationship (in base ten) between adjacent unit types. The place value and face value are combined to determine the value of the digit. We use the context of base five to allow PSTs to work in a system in which they do not have access to preestablished labels for place values or preestablished procedures that might interfere with or mask their sense making.

We leverage base five to provide an opportunity to compare across bases. This may enable PSTs to see and understand more clearly the structure of place value number systems in general and base ten in particular. Prior researchers indicate that, rather than being trapped within base ten, students may see that base ten is an example of a type of place value number system (Treffers, 1987; Vygotsky, 1962). In particular, we will focus on the mathematical task of multiplying by the base (10_{five}) in base five in order to focus on the multiplicative structure of place value number systems.

In base five, the numeral 10_{five} refers to 1 group of five. To avoid confusing base ten and base five, we do not refer to 10_{five} as ten. Rather, we may call 10_{five} a *long*, following the notation given in Figure 1 (Bennett, Burton, & Nelson, 2011). The difference between ten (written 10) and a long (written 10_{five}) was a source of disequilibrium for the PSTs. This disequilibrium was explored within our task, where the PSTs compared the role of multiplying by 10_{five} in base five with the role of multiplying by 10 in base ten, examining the multiplicative structure of both number systems. We explain this in more detail after the solution to the earlier example problem.

A Solution to the Example 3_{five} x 23_{five}

We now solve the problem posed to the reader above. The problem 3_{five} x 23_{five} can be modeled as repeated addition, creating three copies of 23_{five} (Figure 3a). To group like units we rearrange the base five manipula-

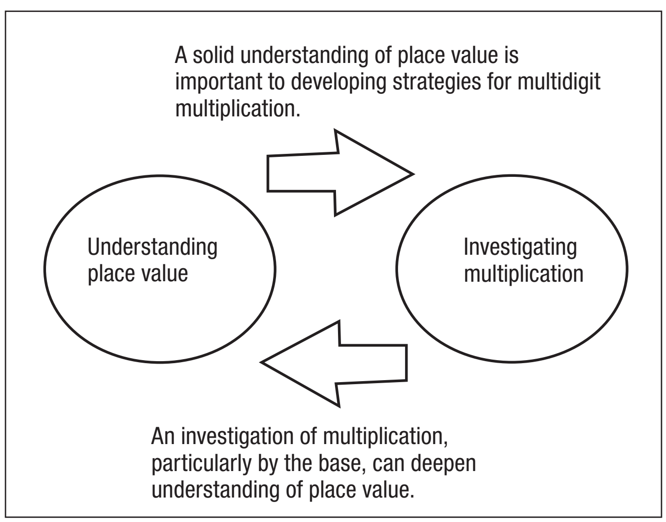


Figure 2. **Linking growth in place value understanding to investigating multiplication.**

tives to place the longs together and to place the ones together (Figure 3b). One characteristic of place value systems is that we need to create minimal collections (i.e., no more than 4 of any one type of unit) for notation, as there are only symbols for values less than the base. To regroup, we look for 5 of one type of unit. Five longs are regrouped to create 1 flat and 5 ones are regrouped to create 1 long (Figure 3c). Once a minimal collection has been established, the solution can then be recorded symbolically as 124_{five} (Figure 3d).

The reader may want to try to find the product of 10_{five} x 23_{five}[1] and reflect on the solution before reading the remainder of this paper. This problem is the central focus of the rest of paper.

Why Times 10?

In base ten, multiplying a number by 10 causes all of the digits in that number to be shifted to the left one place value and a zero to be appended at the end of the number (in the ones place). This is also true in base five when multiplying a number by 10_{five} (or in any base when multiplying by 10_{base}). This phenomenon is important for PSTs to make sense of for two reasons. First, unpacking why digits shift (unchanged) to the left places a focus on the multiplicative structure of base five and base ten place value number systems. Second, multiplication by the base is a key phenomenon to exploit when developing or making sense of multidigit multiplication strategies.[2] Both of these reasons are discussed further in this section, but first we discuss what to name this phenomenon.

What's in a name? There are many options when selecting a name for the multiplication by 10_{five} rule. Possibilities include the *Times Ten Rule, Times One Zero Rule, Times Five Rule*, and *Times Base Rule*. Each of these titles has its affordances and its limitations. In base ten, the Times Ten Rule is an appropriate title, but in base five it is a misleading use of base ten language (Danielson, 2010). The Times One Zero Rule is independent of base, but implies a concatenated digits conception by ignoring place value when naming the number 10_{five}. The Times Five Rule is a name specific to base five rather than being a name that could be applied across bases. Further, five has an associated symbol representation in base ten, namely 5, that is not transferrable to base five. The Times Base Rule is a more mathematically accurate and general title but may not be authentic to the PSTs' own informal language.

For the purpose of this article, we use the Times Base Rule with the caveat that this is not the language that was utilized in the classroom.[3] The Times Base Rule is mathematically accurate, captures the generalizability, and does not have the conceptual limitations of the alternative titles. The Times Base Rule refers to the generalization that multiplying a number by the base (written 10_{base}) of that number system causes the digits of that number to move one place to the left and a zero to be written in the ones place.

Multiplication by 10 and Multiplicative Structure. Multiplication by 10_{base} in any base means that we make *base* copies of that number. In base ten this means 10 copies; in base five this means 5 copies. Having base copies of a unit results in regrouping the *base* copies to 1 copy of the next larger unit. Thus all digits are moved one place to the left in a number and a zero is appended at the end (in the ones place). (See Figure 4.)

In base ten, each digit's place value is 10 times larger than that to its right. For this reason, multiplication by 10 can serve an essential role in probing understanding of

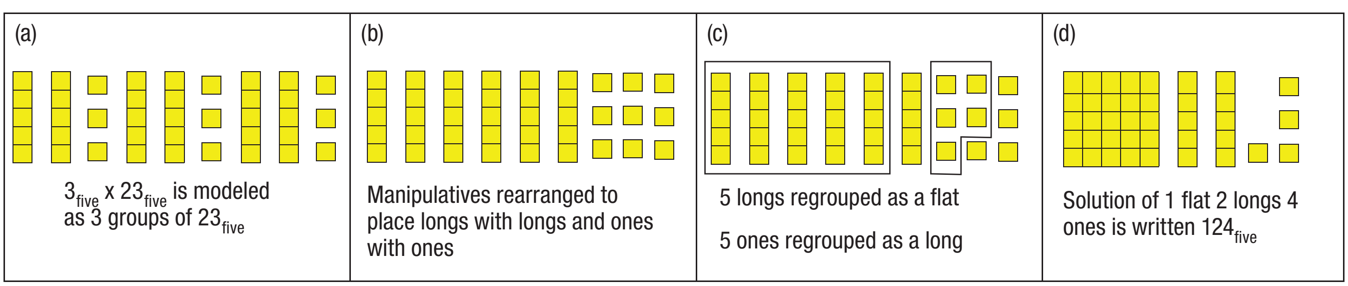


Figure 3. **A solution to 3_{five} x 23_{five}, modeled with manipulatives.**

1 Note: In the task given to PSTs, the numbers in the problem were given in the opposite order, $23_{five} \times 10_{five}$. Nevertheless, the PSTs interpreted the problem as 10_{five} groups of 23_{five} rather than 23_{five} groups of 10_{five}. In the updated version of the task, the prompt is changed to $10_{five} \times 23_{five}$.

2 In fact, this is an essential step in the standard U.S. multiplication algorithm, where numbers are implicitly broken into multiples of powers of ten.

3 The students referred to this rule as the Times Ten Rule or "adding a zero."

The Times Base Rule: Multiplying a number by the base causes all of the digits to shift to the left and a zero to be appended to the end of the number.

(e.g., 10 x 45 = 450 and $10_{five} \times 32_{five} = 320_{five}$)

Note: This rule is true in any base. In this study, PSTs investigate multiplying by 10_{five} in base five.

Figure 4. **Statement of the Times Base Rule.**

the base ten system. Baturo and Cooper (1998) illustrated this point by utilizing *times ten* tasks to show conceptual gaps in students understanding of the multiplicative base ten system. Focusing on multiplication by 10_{base} brings PSTs attention to this consistent multiplicative relationship between unit types (i.e., between columns in multidigit numbers).

Utility of multiplication by 10. Multiplication by 10 and multiples of 10 in base ten also plays an important mathematical role in student generated strategies for multidigit multiplication (Ambrose, Baek, & Carpenter, 2003). Because of the multiplicative structure of the place value system, students often split multiplication problems along place value to make use of the efficiency of multiplying by 10. For example, consider the problem 12 x 23. A student may split the 12 into 10 and 2, changing the problem from 12 groups of 23 to 10 groups of 23 plus 2 more groups of 23. This creates two easier problems (10 x 23 and 2 x 23), the products of which can then be added together. The student generated strategic use of the distributive property may lead to a more general multiplication strategy of splitting numbers along place values.

The role of multiplying by 10 in base ten is essential but is often hidden within the procedure for the standard algorithm. (See Figure 5 for an example of multiplying 23 by 42 using the standard algorithm.) During the first set of steps of the standard algorithm, 3 is multiplied by 42 to result in 126. In the second set of steps, it appears that the digit 2 is multiplied by 42 to get 84. The 2 from 23 is actually 20 (i.e., 10 x 2). Because the value of 2 x 42 is 84, the value of 20 x 42 (which is 10 times more than 2 x 42) can be found by appending a zero to 84 to arrive at 840. Because of the efficiency and compactness of the algorithm, it is possible for students to follow the correct steps and place 84 in the correct location (moved to the left one place) without understanding the relationship between multiplying by 2 and multiplying by 20. Unpacking the meaning of multiplication by 10 may help PSTs to make sense of how multiplication by 20 is related to multiplication by 2. Further, this relationship can serve to highlight the underlying multiplicative structure of place value that allows for the efficiency of the standard algorithm.

Brief Literature Background

Preservice Teachers' Understanding of Place Value, Number, and Operation

A small body of research exists on PSTs' conceptions of number and operation within the traditional base ten context (Thanheiser et al., 2014). A conceptual understanding of multidigit whole numbers requires an understanding of place value and the relevant 10 to 1 relationship found in base ten (Fuson, 1990; Fuson et al., 1997; Kamii, 1986; Thanheiser, 2009a). For example, the 2 in 234 represents 2 groups of a *hundred*. Since a *hundred* is equivalent to 10 *tens* (adjacent unit types), the 2 also represents 20 *tens;* and since a *hundred* is equivalent to 100 *ones* (each *ten* is equivalent to 10 *ones*), the 2 also represents 200 *ones*. PSTs should be able to move flexibly between these three interpretations using the 10 to 1 multiplicative rela-

$$\begin{array}{r} 42 \\ \underline{\times\ 23} \\ 126 \\ \underline{+\ 840} \\ 966 \end{array}$$

Consider the problem 23 x 42. In the standard U.S. algorithm, the multiplicand (42) is multiplied by each digit of the multiplier (23) and then added up with all of the digits properly shifted. The written algorithm is shown on the left. First the 3 is multiplied by the 2, resulting in 6, which is noted below the 3; then the 3 is multiplied by the 4 in 42, resulting in 12, which is noted under the 2, thus resulting in 126 in the first line of the result. Next, the 2 in 23 is multiplied by the 2, resulting in 4, which is shifted one column to the left and noted under the 2 of the 126; then the 2 in 23 is multiplied by the 4, resulting in 8, which is noted to the left of the 4. Sometimes a zero is placed in the ones column after the 84, either to indicate that the 84 refers to 84 tens or merely as a placeholder to remind the algorithm user to shift to the left. Finally, the partial products are summed to arrive at the total of 966.

Figure 5. **A brief description of the standard U.S. algorithm for multidigit multiplication.**

tionships between columns. This flexible understanding of multidigit numbers is referred to by Thanheiser (2009a) as a reference units conception. This conception of number was the most sophisticated held by PSTs in that study, followed by another correct conception in which PSTs were able to see the digits representing the correct values but did not have the flexibility to move between unit types. Thus for the number 234, PSTs would be able to state that the 2 represents 200, the 3 represents 30, and the 4 represents 4, but they would struggle finding alternate interpretations. Thanheiser also identified two incorrect conceptions based on seeing the values of the digits as concatenated. PSTs either saw the digits in 234 as 2 ones, 3 ones, and 4 ones, or combined this incorrect concatenated digits conception with aspects of the correct conception, thus seeing some digits in terms of correct unit types but others incorrectly in terms of ones (i.e., 2 as 2 *hundreds,* 3 as 3 *ones,* and 4 as 4 *ones*). Thanheiser found that two thirds of the PSTs she interviewed held one of the two incorrect conception of multidigit numbers at the beginning of their content (Thanheiser, 2009a) and methods (Thanheiser, 2010, 2014) courses.

PSTs and teachers with an incorrect conception of multidigit numbers struggle to explain the mathematics embedded in standard algorithms (Ball, Hill, & Bass, 2005; Ma, 1999; Thanheiser, 2009a, 2012), such as regrouping, which is often termed *carrying* or *borrowing* in the PSTs' language. Simply informing PSTs about the 10 to 1 relationship between adjacent place values is not sufficient to build conceptual understanding of the structure of the base ten number system (Thanheiser, 2009b, 2014) or to make sense of commonly used multidigit algorithms. Ma's (1999) study of in-service teachers revealed that U.S. teachers were proficient in performing multidigit multiplication, but 61% were unable to provide a conceptually based explanation. The teachers focused on how to perform the algorithm, rather than why the algorithm worked. This is consistent with other research on operations that indicates that PSTs or teachers may believe that knowing how to find an answer is equivalent to understanding (Graeber, 1999; Lo et al., 2008). Despite this documented lack of conceptual understanding, little research has been done to support PSTs' development of place value understanding (Thanheiser, 2014). This study focuses attention on developing an understanding of the multiplicative structure of place value by investigating multiplication by the base in an alternate base context (base five).

Prior Research Using Alternate Bases

Alternate bases have been used as a research and educational tool to both uncover and deepen PSTs' understanding of number and operation. Zazkis and Khoury (1993, 1994) used the context of base five decimal fractions (such as 12.34_{five}) to reveal PSTs' conceptions of the underlying place value structure of numbers that would have been masked by procedural use of base ten language and symbols. The study showed that PSTs struggled to maintain a consistent 5 to 1 multiplicative relationship between place values, sometimes shifting to a 10 to 1 or 2 to 1 relationship instead. The base 20 Mayan numeral system has been used as a context to explore shifts in the value of digits when comparing the values of a "one" with 1, 2, and 6 zeros attached at the end (Thanheiser & Rhoads, 2009; Thanheiser, 2014). This study showed that PSTs struggled to make sense of three-digit and larger numerals within an alternate base.

McClain (2003) and Yackel, Underwood, and Elias (2007) used a base eight context in preservice teacher content courses, focusing on number, coordinating units, and addition and subtraction. Both studies found that learning trajectories from prior research with children were informative in creating learning trajectories for PSTs. The base eight context in those two studies provided an opportunity for PSTs to experience a sense-making approach to early arithmetic. This context provided an alternate view from the procedural approach previously experienced by the PSTs. The work of McClain and Yackel, Underwood, and Elias focused on counting and single and multidigit addition and subtraction of whole numbers. In this study, we build on that work and leverage alternate bases in a multiplication setting. Particularly, we address multiplying by the base as a way to highlight the multiplicative structure of place value number systems.

For this study, our primary research question is: How might PSTs build understanding for multiplication by 10_{base} and how does this impact their understanding of place value structure? In answering this, we (1) show that making sense of multiplication by 10_{base} is not a trivial task, (2) provide insight into how one group of PSTs was able to make sense of multiplication by 10_{five}, and (3) discuss shifts in the PSTs' understanding of place value.

Methods

Description of the Course

The multiplication task discussed below was used in the first quarter of a three-term mathematics content sequence designed for preservice K–8 teachers at a large urban university in the Pacific Northwest. The class met twice a week for 10 weeks, with each class 1 hour and 50 minutes in length. The focus of the first course in the sequence was whole number and operation. The second author was the teacher for the class, while the first author

designed the multiplication tasks and played the role of observer during the class sessions.

In the first five weeks of the term, PSTs investigated number and operation in base ten and alternate bases focusing on counting, addition, and subtraction. The alternate base sequence began with an investigation of a fictional base five context called Alphabitia[4] (Bassarear, 2011), in which PSTs invented and compared number systems using base five manipulatives. The goal of this activity was for the PSTs to develop an understanding of a grouping system and a place value system and the difference between those. In the first two weeks of the term the PSTs also explored historic number systems, including the base 20 Mayan system and the base 60 Babylonian system, as well as the alternate base context of time (hours, minutes, seconds) to examine the underlying structure across the systems (grouping by base, identifying place values, exploring the relationship between place values, etc). Then the PSTs investigated addition and subtraction in base ten and other bases in weeks 3 and 4 and began multiplication in base five in week 5. The central base five multiplication task is shown in Figure 6.

Description of the Task Design

The design of classroom mathematical tasks is a cyclic process of creating, trying, and modifying tasks (Liljedahl, Chernoff, & Zazkis, 2007). This study is part of a larger project that focuses on the design of tasks that leverage alternate bases for supporting PSTs in developing understanding of multidigit multiplication strategies. This article focuses on a portion of one cycle, related to the task of multiplying by the base, and provides insight into how the task was modified for a subsequent cycle.

The design of the initial multiplication task was based on the design heuristic of guided reinvention, taken from Realistic Mathematics Education (RME) (Freudenthal, 1991; Gravemeijer & Terwel, 2000). Students begin in a familiar context where they can use informal strategies to make sense of the situation. Then tasks are provided that move students' mathematical activity to more general and formal strategies. "The idea is to allow learners to come to regard the knowledge they acquire as their own private knowledge, knowledge for which they themselves are responsible" (Gravemeijer & Doorman, 1999, p. 116). By working in base five, the students had the opportunity to reinvent an algorithm for multiplication and develop conceptual knowledge with a degree of ownership. Within this article, we focus on a key stage in the reinvention: making sense of multiplication by the base. In particular, we focus on shifts in PSTs understanding of the structure of place value as they investigated multiplication by the base.

The task was designed to move from simpler to more complex multidigit multiplication problems, with the intention that students would develop more sophisticated strategies based upon the structure of the place value system in order to cope with larger numbers. The multiplication task was designed with the intention of allowing the PSTs to invent their own sense-making solution strategies by placing the PSTs in the nonroutine context of base five rather than base ten. A classroom culture of sharing and explaining ideas was key to the implementation of the task and had been established by the classroom teacher in the prior weeks of class. In the task directions, the PSTs were invited to choose problems to share that would purposefully highlight some aspect of what they learned. The prompt for this task is given in Figure 6. This study

Multiplication Task

Create a poster that demonstrates and explains how to multiply in base 5 (without converting to base 10). Choose two of the following problems to present in your poster, along with an additional problem that highlights something that you find interesting/ challenging/awesome about multiplication in base 5.

Make sure you are showing what you are doing and why. We are interested in the CONCEPT of multiplication. Use the base 5 manipulatives and try to include pictures.

23_{five} x 3_{five}

23_{five} x 10_{five}

42_{five} x 31_{five}

243_{five} x 12_{five}

Figure 6. **The multiplication task given to PSTs.**

4 Alphabitia is similar to the base five investigations in Xmania (Schifter & Fosnot, 1993) and Orpda (Hopkins & Cady, 2007).

focuses on one group of PSTs who chose to investigate the multiplication by 10_{five} subtask.

Description of the Data Collection

The first author observed the two and a half class periods devoted to the multiplication tasks. She took detailed field notes with a focus on the PSTs' strategies and areas of struggle. A videographer captured the work of a focus small group during the first two days of the multiplication study. During the third day, the camera was focused on the whole class discussion rather than the small group. Photographs were taken to capture student work and posters. A brief, written presurvey was collected to provide context for the PSTs' knowledge of the Times Base Rule prior to the start of task (Figure 7).

The small group observed for the study consisted of two women and three men: Alice, Karen, Danny, Lee, and Joe (all names are pseudonyms). One of the five PSTs, Karen, was absent for the second day of the multiplication task.

Description of the Data and Data Analysis

The primary data source was video recordings of small group work within the classroom. In addition, we analyzed responses to a written presurvey (see Figure 7) to determine whether each student demonstrated an awareness of the Times Base Rule in base ten. Rather than focusing on pre- and post-snapshots of knowledge, this study focuses more heavily on the PSTs' learning process, following a recent commendation for more *motion studies of learning* (Thanheiser et al., 2014).

Presurvey

1. Solve the problem 23 x 42 IN TWO DIFFERENT WAYS and explain your reasoning for each step. Include both the how and why for each step, if possible.
2. (a) Show how to solve 24 x 10 and explain your reasoning.

   (b) Show how to solve 24 x 13 and explain your reasoning.

   (c) Are 2(a) and 2(b) the same difficulty level, or is one easier or harder for you? Explain why.

Figure 7. **Prior knowledge about multiplication and Times Base Rule in base ten.**

The focus small group video was analyzed retrospectively with the following steps adapted from Lesh and Lehrer (2000). (1) Each video was watched by the first author, who created a detailed log of events, marking critical events (Maher & Martino, 1996), which provide insight into the generation of student strategies. Critical events included episodes of struggle, discussions of procedures, teacher presses, explanations or justifications within the small group, and comparisons between base five and base ten. For example, one critical event occurred when students misrepresented the number 10_{five} as ten groups rather than 5 groups. (2) Episodes containing critical events were transcribed. (3) The detailed log and transcribed episodes were examined for themes within the base five multiplication data. Themes were identified if they appeared as a common idea or thread linking a number of critical events. Examples of themes include making sense of multiplication by 10_{five}, reliance on procedural knowledge, and comparing base five and base ten. The first of these themes, making sense of multiplication by 10_{five}, is the focal theme of this paper. (4) The detailed video log was reanalyzed to create a narrative of all events related to multiplication by 10_{five}, including transcripts of student conversations and analyses of those transcripts. (5) This narrative was analyzed to find key points in the learning sequence, allowing for the identification of both areas of struggle and conjectures for overcoming struggle.

Results and Discussion

The results section is structured in four parts. These parts focus both on how the PSTs made sense of multiplication by the base and on the PSTs' shifting conceptions of place value structure.

1. We provide brief background information on PSTs' prior engagement with multiplication by 10 in base ten.
2. We show evidence that PSTs struggle to provide a conceptual justification for the Times Base Rule in base five. Their struggles demonstrate a lack of understanding of the multiplicative structure of the place value system.
3. We tell the story of how the PSTs built a conceptual justification for the Times Base Rule by shifting toward viewing 10_{five} as a quantity rather than as two concatenated digits.
4. We highlight how both repeated addition and area imagery were used by the PSTs to develop a better understanding of the Times Base Rule and the multiplicative structure of the place value system.

Background About Small Group's Prior Knowledge

By considering the baseline surveys, we provide a brief snapshot of the PSTs' general awareness of the Times Base Rule in base ten. When comparing 24 x 10 and 24 x 13, three PSTs in the group (Karen, Lee, and Joe) appeared to be aware of the Times Base Rule. The remaining two PSTs (Danny and Alice) did not use the Times Base Rule when comparing the difficulty of the problems. (See Table 1 for the PSTs' responses.) Because several of the PSTs already used the Times Base Rule in base ten, they were primed to notice the same rule occurring during the base five task.

The small group's initial conceptions were similar to the whole class, where 12 of the 20 PSTs used the Times Base Rule and the other eight did not. While the task prompt for the baseline data included a request for explaining their reasoning, the PSTs did not show evidence that they could explain why the Times Base Rule worked. If any explanation was attempted, it tended to involve statements of the following form: 24 x 10 = 240 "because we can move the 0 from the 10 to the 24," using the Times Base Rule as the justification rather than justifying the Times Base Rule.

Struggling to Justify *Times Base Rule* (Procedural Justifications)

During the two days of small group work, the PSTs were asked to attempt four specific base five problems (Figure 5), and they were invited to try out multiplying other numbers to investigate their ideas. Despite teacher encouragement for sense making and an opportunity for inventing algorithms, the small group's strategies initially relied heavily on the standard U.S. algorithm for multiplication, which the PSTs borrowed from base ten and

Table 1
Analysis of each student's level of awareness of a Times Base Rule

PST	Use of Times Base Rule	Evidence, when comparing 24 x 10 to 24 x 13.
Karen	Uses Times Base Rule	"When multiplying a number by 1, 10, or 100 . . . simply add as many or as little zeros as the 1, 10, 100 . . . has." (ellipses were included by the student)
Lee	Uses Times Base Rule	"[24 x 10] is easier because it is a multiple of ten number. I just take the zero from ten and put it to the right of 24 for shortcut math in my head."
Joe	Uses Times Base Rule	24 x 10 = 240 = 312 24 x 3 = 72 "You can break multiplication down to easier problems. I broke it into base 10 then the remainder 3, then added." "[24 x 10] is easier [than 24 x 13]. When big #'s are multiplied by more than 10 it is hard for me and I have to break it down to groups of 10."
Alice	No evidence of use of Times Base Rule	"They are similar for me. If you think of it as groups, it is the same."
Danny	No evidence of use of Times Base Rule	a) 24 x 10 00 +240 240 b) 1 24 x 13 1 72 + 240 312 "[24 x 10] is less difficult than [24 x 13] as it is multiplied by 10, [24 x 10] makes for an easier process and the problem does not include any carrying."

modified to work in base five. (An example of the standard U.S. algorithm for multiplication was given in Figure 5 and its modification for base five is given in Figure 8a.)

The group's first solution for 23_{five} x 10_{five} relied entirely on the standard algorithm, which was written out to include a row of zeros, followed by 230, and summed to 230 (Figure 8b). The teacher pressed the PSTs to make sense of their solution. The following episode demonstrates the PSTs' reliance on the standard algorithm to justify why the Times Base Rule works for this example.

Teacher: So why can we just add a zero to the end?

Joe: Because if you do it the standard algorithm way it works and that's why I do that.

Karen: When you multiply by 1 you always know it's going to be the same number. So when it's a ten or hundred or a thousand[5] you add as many zeros as are behind that number.

Teacher: Why does that work? Why do you always add a zero there?

Joe: Because you put a one, a ones column there, and the zeros are just place holders.

The PSTs' explanations treat the algorithm as the source of justification. The standard multiplication algorithm treats each digit in a multidigit problem as a single digit independent of place value. The efficiency of the algorithm allows its users to solve large problems knowing only single digit facts. Perhaps it is this aspect of the algorithm that invites the PSTs to see the 0 and 1 in 10_{five} as two separate issues to explain, rather than looking at multiplication by 10_{five} as a number of copies of a specified number. This can be seen in Karen's statement above about multiplying by 1 causing a number to remain the same. Continuing the conversation, Karen deals with the zero separately.

Karen: When you multiply things by zero it's zero. So I think that's why we're getting that whole "place holder thing" [student gestures air quotes]. 'Cause it's just adding zeros. When you go down the line you have to add more zeros anyway. [She makes hand gestures as if she's writing out the algorithm, pointing at the zeros.]

While the group has developed a type of procedural explanation that relies on the standard algorithm, they continue to express discomfort with the explanation. We see evidence in the conversation that follows Karen's previous explanation.

(a) 1 42_{five} x 23_{five} 231 +134 2121_{five}	In order to modify the standard algorithm to base five, PSTs regroup whenever they have 5 or more of something. For example, 3 ones x 2 ones = 6 units, which is regrouped as 1 long 1 ones, with the 1 long "carried" above the 4 and added to the product of 3 units x 4 longs. This results in thirteen longs, which is, in turn, regrouped to 2 flats 3 longs. It is unknown whether the PSTs consider the values as units, longs, and flats, or merely concatenated digits.
(b) 23_{five} x 10_{five} 00 +23 230_{five}	The steps of the standard algorithm are dutifully followed, resulting in a row of 0s when 0 is multiplied by 23_{five}. In the second step, the 1 in 10_{five} is multiplied by 23_{five}, resulting in a 23 shifted one place to the left. The 00 is added to the 23 to arrive at 230_{five}. The separate steps of the algorithm were used as a justification for why multiplication by 10_{five} results in adding a zero to the end of a number.

Figure 8. **The Standard U.S. Algorithm for multiplication in base ten and modified to base five.**

5 Karen has incorrect use of base ten language in a base five context.

Danny: Is there another way to describe it, though? The one copies the number, the zeros just add. It's really easy to think of. And it works in base five.

Karen: You know if you just say it, a kid will be like, Oh that's what I have to do . . . but why? It's hard to explain.

Danny: See I can't even remember, I can't explain why in regular base ten.

We see students looking back to their base ten knowledge and actively making connections between the base ten and base five systems. What they know in one system, they may import to the other system (e.g., standard algorithm), while what they don't know (e.g., explanations) may become more apparent to both the teacher/researcher and the PSTs. Danny's acknowledgment that he cannot explain why the Times Base Rule works in base ten may demonstrate that he is reflecting on his own conceptual knowledge gaps as a result of investigating within base five.

Shifting from 10_{five} as Two Distinct Digits to 10_{five} as a Quantity

After a teacher suggestion that the group tries to unpack the Times Base Rule to make it more accessible to children, Karen returns to the group's bag of manipulatives and takes out 2 longs, 3 units, and 1 long, creating 23_{five} and 10_{five}. While Karen's selection of manipulatives (Figure 9a) represents both numbers in the base five problem, she struggles to use accurate language in describing the numbers.

Karen: If it's twenty times ten,[6] there's ten sets of twenty-three.

Joe: Yeah

Karen: This is going to take a while.

Karen's use of base ten language to talk about base five is problematic. Multiple group members initially make the mistake of gathering ten [base ten] copies of 23_{five} instead of gathering five (10_{five}) copies.

The small group works through this confusion by inventing a strategy based on repeated addition, which Danny voices below (visual provided in Figure 9b.)

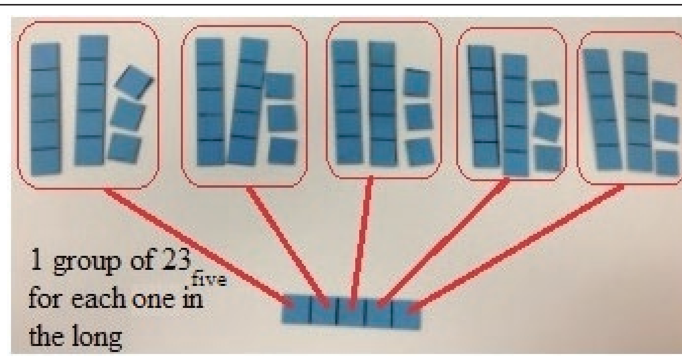
Danny: If we could line this [pointing to 1 long] up with the groups we are using, there would only be one, two, three, four, five groups. [As he talks he points to one unit at a time on the long, until he gets to the fifth unit.]

Joe: Of two longs and three units.

Joe: Yeah, I like how you are touching it. [Referring to Danny counting out on the long manipulative to find the number of groups of 23_{five}.]

The process of touch counting with the long (10_{five}) manipulative allows the group to unpack the meaning

(a) Karen represents 23_{five} with 2 longs 3 ones, and she represents 10_{five} with 1 long. Verbally, she misnames the 10_{five} as ten.



(b) Danny associates each single square in the long with 1 group of 2 longs 3 ones, allowing him to make explicit use of both numbers in the product.

Figure 9. **Using manipulatives to visualize $23_{five} \times 10_{five}$.**

6 Karen refers to $20_{five} \times 10_{five}$ but uses base ten language inappropriately, calling 20_{five} twenty and 10_{five} ten.

of 10_{five} both as 1 long and as 5 units. This represents a shift from the previous strategy of treating 10_{five} as two concatenated digits, "1" and "0." The strategy also allows the group to overcome the confusion between 10_{ten} and 10_{five}, differentiating between the meaning in base ten and the meaning in base five.

Multiplication by 10_{five}: Assembling a Conceptual Justification

The small group discussion has shifted from treating 10_{five} as a 1 and a 0 and is now treating 10_{five} as a quantity that can be considered either as 1 long or as 5 units. With this conception in mind, the teacher presses once again for making sense of the Times Base Rule. This time the teacher prompt shifts from asking why a zero is placed at the end of a number and focuses on why the rest of the digits have shifted (without change) to the left.

Teacher: So what happens when you move it over? What does that mean?

Alice: It's holding a place or something. The fact that there's nothing in that column.

Teacher: So what caused it to move over?

Joe: Because we have enough.

Alice: We had enough to go to the next column.

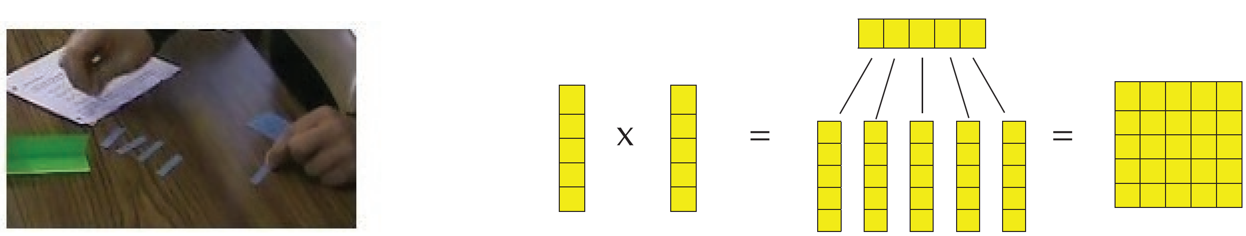
Joe: So in this case you have five groups of this which bumps it over to the next group and you have five groups of this which bumps it over to the next group. Because that's what it takes, the qualification to move over to another group is to completely be full with five units of that, so if you multiply by five you have five units of this which puts you over here [gesturing at an empty table] the same amount of five, just, you know, five units of it, so it goes in the next one and then you have five units of this thing so that goes to the next one.

This new explanation is not based on the standard algorithm, but instead relies on the consistent multiplicative structure of the place value system. The PSTs' language highlights the consistent 5 to 1 multiplicative relationship between adjacent unit types as a result of investigating the Times Base Rule. This can be seen in Joe's language about having five groups and in his explicit attention to the number of items needed to allow for regrouping. Next Joe explains why a long times a long is a flat.

Joe: For every one block here [pointing to each unit on the long] you get one of these [points to a long]. So you get one is one [associates a long for each unit he counts on the original long] two equals two, three equals three, four four, a fifth one gets you up to these [pointing to a flat] so if you were to do that for three [pointing to 3 longs], you'd just get three of these [flats].

The PSTs assemble two key pieces of information to create this conceptual explanation.

1. Multiplication by 10_{five} means repeatedly add a number 5 times.



For each unit he touches in the first long (the multiplier), Joe counts out another long. This results in a collection of 5 longs, which he regroups to 1 flat.

Figure 10. **Joe devises a touch counting strategy for multiplying 10_{five} x 10_{five}.**

2. Adding 5 copies of one unit type allows you to upgrade to the next larger unit type.

While students may have held the two pieces of this idea separately, combining them into a conceptual explanation for the Times Base Rule occurred only after continued teacher press for justification. The conceptual explanation created by the students' touches on the very structure of the base five number system: that each column or place value represents a grouping of 5 times the previous column. This explanation stands in contrast to the earlier procedural explanations that treated the 1 and 0 in 10_{five} as two separate digits. This indicates a shift from discussing concatenated digits toward explaining the multiplicative structure of a place value system.

Multiplication by 10_{five}: Making Generalizations & Comparing Strategies

After using the touch counting method to justify why a long times a long is a flat, the small group begins to generalize this strategy.

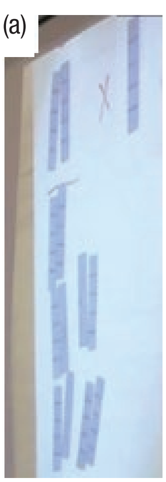
Joe: So you used to have three of these [longs] living is this column and now you have three of these [flats]. So you just move over a column. And if you had these [flats] you'd move over to a different column.

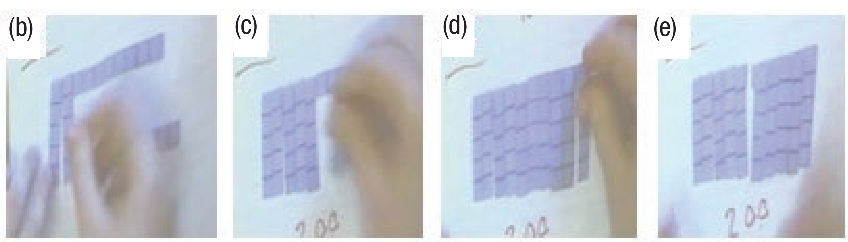
Mathematically, Joe uses the fact that 10_{five} x 10_{five} $=100_{five}$ to conclude that 10_{five} x 30_{five} = 300_{five} and that 10_{five} x 100_{five} = 1000_{five}. Thus, it seems that Joe is explicitly aware that if a base five manipulative (or quantity of matching manipulatives) is multiplied by a long, it upgrades to the next larger type of base five manipulative.

Near the end of the second class period, each small group is asked to share their investigations with the whole class. Members of the focus small group share their exploration of multiplication by 10_{five} at the board, focusing on the touch counting approach for repeated addition using the example 20_{five} x 10_{five} (Figure 11a). Viveca, a member of a different small group, requests permission to rearrange the group's manipulatives. Viveca sets the two factors perpendicular to each other (Figure 11b–e) and creates a rectangular array of longs. The class spontaneously applauds this new area model, and Danny announces to his group: "That was a breakthrough." Having created a justification for the Times Base Rule through their repeated addition model of multiplication, they were primed to make sense of an area based justification for the Times Base Rule. This justification creates a visual model that focuses on the multiplicative structure of the base five system. The number being multiplied by the base (in Figure 11, this is 20_{five}) is treated as one dimension of the rectangle, while the base is treated as the other dimension. The visual model highlights how each long is multiplied by 10_{five} and becomes a flat. The structure of the number system, as well as the area model imagery, continued to be a focus of later class discussions of multidigit multiplication in both base five and base ten.

Overview of the Shifts in Place Value Understanding

When investigating why the Times Base Rule would make sense in base five, the small group initially focused on the separate digits of 1 and 0 in 10_{five}, pointing to the





(a) The focus small group presented 20_{five} x 10_{five} to the whole class by creating one group of 2 longs for each one in 10_{five} (1 long).

(b) Viveca (from another small group) rearranges the manipulatives for 20_{five} x 10_{five} to create a rectangle that is 2 longs by 1 long.

(c)–(e) Viveca fills the rectangle with longs, using the area model to find the solution of 2 flats, or 200_{five}.

Figure 11. **Rearranging the manipulatives to visualize the area model.**

separate ideas that multiplication by 1 causes no change to the other factor, while multiplication by 0 results in a zero. These two facts were assembled haphazardly as a justification for why a zero was placed at the end of a number when multiplying by 10_{five}. An important shift in understanding of the Times Base Rule appears to be related to reconceiving of 10_{five} as a quantity (equivalent to 5 units) rather than as two separate single digits.

Using the language of Thanheiser's framework (2009a) for conceptions of multidigit numbers, the PSTs seem to have shifted from one of the concatenated digits approaches toward a groups of ones and/or reference units approach. Moreover, the PSTs recognized that 10_{five} happens to be the special quantity that results in regrouping. That is, if you add 10_{five} copies of a single item you get one copy of the next larger item. For example, 10_{five} copies of a long results in a flat. The importance of conceiving of 10_{five} as a quantity is consistent with research by Kamii (1986), who notes a similar importance for children when thinking of ten.

Reflections

Modifications to the Task

The data presented in this study represents part of one cycle in the development of a classroom mathematical task (Liljedahl et al., 2007). Each cycle consists of *design, enact, reflect,* and *modify/redesign* phases. The last part informs changes to the task for future cycles. Two key changes will be made in the task for the next cycle. First, rather than giving the task as four subtasks, the subtasks will be given one at a time, to allow for classroom discussions and cross-pollination of ideas between subtasks. While the focal group for this study was able to make sense of multiplication by the base, the other small groups did not devote as much attention to the Times Base Rule and may have missed an opportunity to develop a better understanding of the multiplicative structure of place value systems. By breaking the task into four subtasks, all small groups would be encouraged to unpack the Times Base Rule. The Times Base Rule is an important phenomenon to investigate because it serves as a building block for making sense of multidigit multiplication as well as a tool for making sense of place value structure.

The second overall task change that will be made for the next cycle is a stronger focus on the guided reinvention of a general strategy for multidigit multiplication. (See Figure 6 for the prior task sequence and Figure 12 for the modified task sequence.) The PSTs will be asked to create a general strategy for finding the area of any rectangle in a box of base five rectangles. We anticipate that PSTs will make use of the Times Base Rule and the multiplicative structure of base five when reinventing their own multiplication strategies.

This refined task sequence will be subject to further study and revision in the next iteration of the task design study. Attention will be paid to prompts and investigations that promote PSTs' guided reinvention of multidigit multiplication strategies.

The Role of an Alternate Base

The exploration of base five as a context for number and operation has two intended pedagogical goals. The first goal is to provide an alternate base to compare against base ten. The differences between the bases can serve to draw attention to the meaning of place value and the relationships between columns. The similarities between bases, particularly the Times Base Rule, can draw attention to the multiplicative structure of place value number systems. In alignment with the Common Core Mathemati-

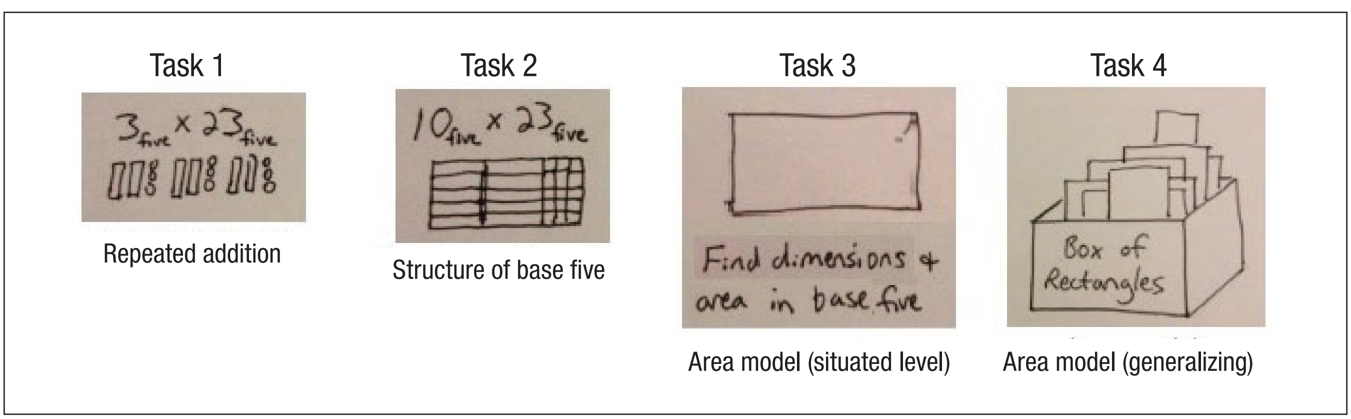


Figure 12. **Modified instructional sequence for multidigit multiplication in base five.**

cal Practice *Look for and make use of structure* (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), the structures of base five and base ten become a focal point of class discussion and student reasoning.

The second goal in using base five was to provide PSTs with the type of sense-making learning experience advocated by the current reform movement and by the tenets of RME. This was done by placing the PSTs in a context where they might explore ideas in number and operation without relying on their prior wealth (or burden) of memorized number facts and rules. This opportunity was successful in leading PSTs to focus on sense making and the structure in mathematics, rather than on memorizing and practicing rules. While base five was a convenient base to use for this teaching experiment (it aligned with the course textbook), we imagine that investigating other alternate bases could also serve the purposes of (a) providing a comparison to base ten and (b) providing a context for sense-making experiences. The discussion of which base is best is a discussion we will leave for the readers.

Take-away

While PSTs may be aware that when you multiply by 10 you can just append a zero, they may not be able to justify why this is true. Investigating the Times Base Rule in alternate bases serves to bring forward the question of why this phenomenon occurs generally and what that indicates about the structure of place value number systems. Creating a valid justification for the Times Base Rule is not a trivial task for PSTs, but it is a worthy task for two mathematical reasons. First, making sense of the Times Base Rule helps PSTs develop a conceptual understanding of the multiplicative structure of place value number systems. Second, the Times Base Rule is a fundamental part of making sense of multidigit multiplication strategies, including both student-invented strategies and standardized algorithms.

References

Ambrose, R., Baek, J. M., & Carpenter, T. P. (2003). Children's invention of multidigit multiplication and division algorithms. In A. J. Baroody & A. Dowker (Eds.), *The Development of Arithmetic Concepts and Skills: Constructive Adaptive Expertise* (pp. 307–338). Mahwah, NJ: Taylor & Francis.

Ball, D. L. (1988). *Knowledge and reasoning in mathematical pedagogy: Examining what prospective teachers bring to teacher education* (Doctoral dissertation). Michigan State University. Retrieved from http://www-personal.umich.edu/~dball/books/DBall_dissertation.pdf

Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, *29*(1), 14–17, 20–22, 43–46.

Bassarear, T. (2011). *Mathematics for elementary school teachers* (5th ed.). Belmont, CA: Brooks/Cole, Cengage Learning.

Baturo, A. R., & Cooper, T. J. (1998). Construction of multiplicative abstract schema for decimal-number numeration. In *Proceeding of the 23rd Conference of the International Group for the Psychology of Mathematics Education*. Retrieved from http://eprints.qut.edu.au/31380/

Bennett, A., Burton, L., & Nelson, T. (2011). *Mathematics for elementary teachers: A conceptual approach* (9th ed.). New York: McGraw-Hill Science/Engineering/Math.

Danielson, C. (2010). What is ten? Relationship between language and numeration. *For the Learning of Mathematics*, *30*(1), 23–25.

Freudenthal, H. (1991). Revisiting mathematics education: China lectures. Norwell, MA: Kluwer Academic Publishers. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED341576

Fuson, K. C. (1990). Conceptual structures for multiunit numbers: Implications for learning and teaching multidigit addition, subtraction, and place value. *Cognition and Instruction*, *7*(4), 343–403.

Fuson, K. C., Wearne, D., Hiebert, J. C., Murray, H. G., Human, P. G., Olivier, A. I., . . . Fennema, E. (1997). Children's conceptual structures for multidigit numbers and methods of multidigit addition and subtraction. *Journal for Research in Mathematics Education*, *28*(2),130–162.

Graeber, A. O. (1999). Forms of knowing mathematics: What preservice teachers should learn. *Educational Studies in Mathematics*, *38*(1/3), 189–208.

Gravemeijer, K., & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics*, *39*(1/3), 111–129.

Gravemeijer, K., & Terwel, J. (2000). Hans Freudenthal: A mathematician on didactics and curriculum theory. *Journal of Curriculum Studies*, *32*(6), 777–796.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hopkins, T. M., & Cady, J. A. (2007). What is the value of @*#? Deepening teachers' understanding of place value. *Teaching Children Mathematics*, *13*(8), 434–437.

Kamii, C. (1986). Place value: An explanation of its difficulty and educational implications for the primary grades. *Journal of Research in Childhood Education*, *1*(2), 75–86.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Lesh, R., & Lehrer, R. (2000). Iterative refinement cycles for videotape analyses of conceptual change. In R. Lesh & R. Lehrer (Eds.), *Handbook of Research Design in Mathematics and Science Education* (pp. 665–708). Mahwah, NJ: Lawrence Erlbaum Associates.

Liljedahl, P., Chernoff, E., & Zazkis, R. (2007). Interweaving mathematics and pedagogy in task design: A tale of one task. *Journal of Mathematics Teacher Education*, *10*(4/6), 239–249.

Lo, J.-J., Grant, T. J., & Flowers, J. (2008). Challenges in deepening prospective teachers' understanding of multiplication through justification. *Journal of Mathematics Teacher Education*, *11*(1), 5–22.

Ma, L. (1999). *Knowing and teaching elementary mathematics teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum Associates.

Maher, C. A., & Martino, A. M. (1996). The development of the idea of mathematical proof: A 5-year case study. *Journal for Research in Mathematics Education*, *27*, 194–214.

McClain, K. (2003). Supporting preservice teachers' understanding of place value and multidigit arithmetic. *Mathematical Thinking & Learning*, *5*(4), 281–306.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA.

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

Schifter, D., & Fosnot, C. T. (1993). Reconstructing mathematics education: Stories of teachers meeting the challenge of reform. Retrieved from http://eric.ed.gov/?id=ED387357

Thanheiser, E. (2009a). Preservice elementary school teachers' conceptions of multidigit whole numbers. *Journal for Research in Mathematics Education*, *40*(3), 251–281.

Thanheiser, E. (2009b). Understanding base-ten numbers: Why don't you just tell them it is ten? In S. L. Swars, D. W. Stinson, & S. Lemons-Smith (Eds.), *Proceedings of the 31st Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1292–1298). Atlanta, GA: Georgia State University.

Thanheiser, E. (2010). Investigating further preservice teachers' conceptions of multidigit whole numbers: Refining a framework. *Educational Studies in Mathematics*, *75*(3), 241–251.

Thanheiser, E. (2012). Understanding multidigit whole numbers: The role of knowledge components, connections, and context in understanding regrouping 3+-digit numbers. *Journal of Mathematical Behavior*, *31*(2), 220–234.

Thanheiser, E. (2014). Developing preservice elementary teachers conceptions with well-designed tasks: Explaining successes and analyzing conceptual difficulties. *Journal of Mathematics Teacher Education*. doi:10.1007/s10857-014-9279-9

Thanheiser, E., Browning, C., Edson, A. J., Lo, J.-J., Morton, C., Whitacre, I., & Olanoff, D. (2014). Mathematical content knowledge for teaching elementary mathematics: What do we know, what do we not know, and where do we go from here? *Montana Mathematics Enthusiast*, *11*(2), 433–448.

Thanheiser, E., Philipp, R. A., Fasteen, J., Strand, K., & Mills, B. (2013). Preservice-teacher interviews: A tool for motivating mathematics learning. *Mathematics Teacher Educator*, *1*(2), 137.

Thanheiser, E., & Rhoads, K. E. (2009). Exploring preservice teachers' conceptions of numbers via the Mayan number system. In S. L. Swars, D. W. Stinson, & S. Lemons-Smith (Eds.), *Proceedings of the 31st Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1220–1227). Atlanta, GA: Georgia State University.

Treffers, A. (1987). *Three dimensions: A model of goal and theory description in mathematics instruction—The Wiskobas project*. Dordrecht, Holland: Reidel Publishing Company. Retrieved from http://www.nrdc.org.uk/text.asp?ref=%2Fanr_details.asp%3FID%3D2071

Vygotsky, L. S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.

Yackel, E., Underwood, D., & Elias, N. (2007). Mathematical tasks designed to foster a reconceptualized view of early arithmetic. *Journal of Mathematics Teacher Education*, *10*(4-6), 351–367.

Zazkis, R., & Khoury, H. A. (1993). Place value and rational number representations: Problem solving in the unfamiliar domain of non-decimals. *Focus on Learning Problems in Mathematics*, *15*(1), 38–51.

Zazkis, R., & Khoury, H. A. (1994). To the right of the decimal point: Preservice teachers' concepts of place value and multidigit structures. *Research in Collegiate Mathematics Education*, *1*, 195–224.

Authors

Jodi Fasteen, Neuberger Hall, Room 326,
724 SW Harrison Street, Portland, OR 97207-0751;
jfasteen@pdx.edu

Kathleen Melhuish, Neuberger Hall, Room M428,
724 SW Harrison Street, Portland, OR 97207-0751;
melhuish@pdx.edu

Eva Thanheiser, Neuberger Hall, Room 323,
724 SW Harrison Street, Portland, OR 97207-0751;
evat@pdx.edu

Elementary Teacher Candidates' Use of Number Strings: Creating a Math-Talk Learning Community*

Laura Bofferding
Purdue University

Melissa Kemmerle
Stanford University

This article presents the results of an exploratory study detailing 4 teacher candidates' initial implementations of a number string protocol in which they presented sequences of related problems to 3rd graders. We detail how the teacher candidates were taught the components of the protocol in their methods course and describe the math-talk (student-participation) levels that occurred during their 1st number string experience with their students. We coded the lesson transcripts for math-talk levels, which range from teacher-led to student-driven, and provide examples of the number strings and excerpts from the teacher candidates' reflections to illustrate our results. Results indicate that number strings are a supportive structure for beginning teachers as they facilitate math talk.

Key words: Elementary teacher candidates; Math-talk levels; Number strings; Sequences of related problems

With the introduction of the Common Core State Standards for Mathematical Practice, there is renewed momentum for mathematical discussion in classrooms. Teachers are expected to elicit students' reasoning and explanations as well as help students listen and respond to each other's mathematical thinking (National Council of Teachers of Mathematics, 2000; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). These processes are difficult to manage, especially for new teachers who often experienced teaching as telling (Smith, 1996; Nicol, 1999): "There is a strong tendency for novice teachers, once they have entered the profession, to revert to their default model of teaching as they were taught" (Borg, 2004, p. 275). Further, teachers may rely on "telling" to avoid the daunting task of making sense of students' thinking (Smith, 1996). Without purposeful planning about how to respond to students' comments in a mathematical discussion, new teachers are often not able to give pedagogically meaningful responses to their students and often ignore unanticipated responses, hindering the meaning-making process (Inoue & Buczynski, 2011).

As teacher educators, one of our goals is to help teacher candidates move beyond telling to eliciting and building on students' thinking. One way we can do this is through modeling a specific routine, during which teacher candidates participate as students in the lesson and we as teacher educators periodically step outside our elementary teacher roles to discuss the elements of the practice. Along with this, we can provide classroom and clinical experiences where teacher candidates are directly involved in rehearsing this particular routine and reflecting on their practice (Grossman et al., 2009). In this article, we present results from an exploratory study on teacher candidates' use of a specific routine or practice—number strings—and explore its potential to support novice teachers' efforts to facilitate mathematics discussions.

Number Strings

Researchers have identified number strings[1] as a discussion routine that could help teachers focus less on *teaching as telling* and more on *exploring* the mathematics and *facilitating* conversations about students' solutions and reasoning (Lampert, Beasley, Ghousseini, Kazemi, & Franke, 2010). Number strings consist of a series of related problems that highlight a particular mental math strategy or big idea. Each problem in a particular string is written horizontally and is completed one at a time; students mentally work out a solution to the problem and share their methods with the class (DiBrienza & Shevell, 1998; Fosnot & Dolk, 2001; Parrish, 2011). For example, students might first see 3 x 4, then 3 x 8, then 6 x 8, followed by 6 x 16. The underlying strategy here is doubling, although students might also use other mental

1 It may be more appropriate to call this routine *problem strings*; however, we use the term *number strings* to reflect popular terminology. For a more detailed account of number strings and how they differ from other forms of math talk, see numberstrings.com/2014/05/27/whats-in-a-name/.

*This manuscript is based on a presentation given at the 2014 annual meeting of the Association of Mathematics Teacher Educators in Irvine, CA.

math strategies. When conducting the number string routine, teachers are encouraged to elicit and probe student strategies, record these strategies to facilitate clarity, foster an understanding of representations, encourage comparisons, address incorrect answers, and encourage students and their classmates to clarify or justify their strategies, especially in terms of efficiency and utility (Lampert et al., 2010).

There are a few accounts of teachers implementing number strings with their students (e.g., Fosnot & Dolk, 2001; Lampert et al., 2010; Parrish, 2011), and some include a focus on the types of strategies that students develop when given the opportunity (e.g., Murata et al., under review; O'Loughlin, 2007). However, there is little research on how teacher candidates implement the routine and whether those who use them can do so faithfully and maintain higher levels of math talk with elementary students. In the following section, we describe a framework for characterizing levels of math talk and discuss how the elements embedded in the number string routine have the potential to support higher levels of math talk.

Math-Talk Theoretical Framework

Hufferd-Ackles, Fuson, and Sherin (2004) illustrated four main components that contribute to a math-talk learning community: questioning, explaining mathematical thinking, source of mathematical ideas, and responsibility for learning. They describe each of these components within a series of levels, highlighting a trajectory toward effectively incorporating many of the Common Core mathematical practices into teaching. Scoring a zero across all four math-talk components constitutes the lowest level, at which teaching involves telling, asking right/wrong questions, dictating how to solve problems, and making little attempt to elicit student thinking. Math talk at levels 1 and 2 involves a shift in focus from answers and teacher-provided explanations to students providing explanations of their solution strategies. The teacher begins to ask probing questions and involve other students in helping make sense of each other's strategies. As teachers and students shift toward using discourse to "extend one's own thinking as well as the thinking of others" (p. 82), they advance to the highest level within the framework and a more learner-centered mathematics community, a level 3.

Questioning

Teachers may ask students questions in order to clarify terminology, lead students through a method, extend students' thinking, and involve other students in the discussion (Boaler & Brodie, 2004). Further, the style, substance, and quantity of teachers' questions greatly affect the classroom learning environment (Kemmerle, 2013; Hiebert & Wearne, 1993). In low math-talk classrooms, teachers do not leverage questioning; instead, they ask yes/no or quick computation questions (Hufferd-Ackles et al., 2004). Teacher candidates struggle with asking questions in order to learn more about students' thinking; instead, they often ask questions to move students toward the correct answer (Nicol, 1999). When students fail to answer a question or provide an incorrect answer, teacher candidates might ask leading questions or have difficulty responding (Inoue & Buczynski, 2011). Because one of the first steps in the number string routine involves additional questioning, our conjecture was that teacher candidates using the routine would move beyond telling and answer seeking; using the routine, teachers would elicit and encourage students to explain their strategies by using a low-press question (Kazemi & Stipek, 2001), such as "How did you get that?" Teachers' strategy-eliciting exposes students to a diversity of methods and helps them progress toward thinking about when particular methods are more efficient (National Research Council, 2001). Further, part of the number string protocol (see Kazemi, Franke, & Lampert, 2009, for details) involves asking additional low-press, clarifying questions (Chapin, O'Connor, & Anderson, 2009), such as "Where does the 2 come from?" or "How did you decompose the 12?" as well as some high-press questions that require more justification (Kazemi & Stipek, 2001), such as "Why did you add instead of subtract?" A final aspect of number strings that encourages high-press questions and elevates the level of math talk is having students consider connections among mathematical ideas, strategies, and representations (DiBrienza & Shevell, 1998; Kazemi & Stipek, 2001; National Research Council, 2001; Parrish, 2011). Teachers can do this by asking questions such as "How are these two strategies similar or different?" or "How can we change our representation to show the new strategy?"

Explaining Mathematical Thinking

Through questioning, teachers can elevate the class's mathematical discourse by encouraging students to explain their solution strategies. However, teacher candidates often define successful teaching as providing clear explanations (Minor, Onwuegbuzie, Witcher, & James, 2002) and may feel compelled to explain mathematical concepts and procedures in order to meet their perceptions of what teachers do (Parrish, 2011). This practice results in low math-talk discourse (Hufferd-Ackles et al., 2004). As teacher candidates shift toward letting students lead the explanations, they encounter the challenge of attending to and noticing the relevant details of students' mathematical explanations (Jacobs, Lamb, & Philipp, 2010). Moreover, if teacher candidates do not understand

a student's reasoning, they may lose confidence in their own understanding of mathematics (Nicol, 1999). The number string routine encourages teacher candidates to use revoicing and restating, which help students learn to explain more thoroughly (Chapin et al., 2009). When revoicing, teachers repeat or summarize what students say using more precise language and ask students to evaluate whether the statement accurately represents their thoughts. This practice helps students justify and take ownership of their ideas, and "the revoicing move entails a collaborative effort at building and explicating a complex idea" (O'Connor & Michaels, 2007, p. 281). As more students restate (Chapin et al., 2009) or clarify each other's ideas, the class moves to higher levels of math talk (Hufferd-Ackles et al., 2004).

Source of Mathematical Ideas

Many teacher candidates espouse a view that their role is to dispense information (Brookhart & Freeman, 1992), which can lead to low math-talk classrooms where teachers do the explaining and are the source of mathematical ideas (Hufferd-Ackles et al., 2004). Because teacher candidates also value being student centered (Minor et al., 2002), they may compensate by using "show and tell" (Stein, Engle, Smith, & Hughes, 2008)—having multiple students explain their strategies without providing input, without supporting other learners to learn from the explanation, or without attempting to help them make connections among their strategies. In classrooms with higher levels of math talk, teachers use students' ideas to a greater extent; therefore, students' ideas guide mathematics lessons. Additionally, students' more detailed explanations engender comparisons (Hufferd-Ackles et al., 2004), especially when they are given time to grapple with a variety of strategies (Murata et al., under review).

The number string routine can help teacher candidates transition to building on students' thinking and to giving students more control over the conversation. One important element of the number string routine is accepting all student answers, even incorrect ones, when first posing a problem. This encourages students to be the source of the ideas rather than the teacher. At the same time, teachers can use the incorrect answers as an opportunity for students "to confront their thinking" (Parrish, 2011). Another integral part of the number string routine, representing students' strategies, ensures that students have access to each other's methods and also provides a means for comparing students' strategies visually (Fosnot & Dolk, 2001). Although teacher candidates may have difficulty using representations to highlight similarities and differences *among* problems (Bofferding, 2012), with the visuals as a guide, students can talk about each other's strategies, resulting in higher math talk. The comparisons are important because making connections among strategies and the underlying mathematics is a key purpose of number strings.

One prevalent model for representing student thinking is the empty number line, which encourages multiple strategies (Klein, Beishuizen, & Treffers, 1998), helps highlight potential shortcuts (Bobis, 2007), and can improve students' procedural competence (Klein et al., 1998)—all of which are goals of number strings. Empty number lines best illustrate jumping strategies (adding on or finding a difference) or showing addition and subtraction together, such as when students round one number and compensate for the change at the end (Fosnot & Dolk, 2001). Some strategies, however, such as splitting, do not lend themselves to illustration via the empty number line (Bobis, 2007). Splitting involves decomposing numbers according to their place values and adding the parts in stages. In these situations, the use of branching or recomposing notation (drawing lines from the number to show how it is broken up) is more appropriate (Fosnot & Dolk, 2001). (An example of an empty number line and the use of branching notation can be found in Table 1 for String B.)

Responsibility for Learning

When teachers are the source of mathematical ideas—that is, their classrooms are at a lower level of math talk—they also hold responsibility for the learning of others (Hufferd-Ackles et al., 2004). In an effort to encourage students to take a more active role in conversations, work together to make sense of the mathematics, and feel more comfortable with sharing, teachers can use a think-pair-share strategy or partner talk (Chapin, O'Connor, & Andersen, 2009; Reinhart, 2000). Aside from making all students responsible for solving the problems, talking with a partner could improve students' explanations, making them more confident in sharing. Further, as students participate in a classroom where their thinking is valued, they will begin to take a more active role in their learning (Hufferd-Ackles et al., 2004).

Based on a survey of 134 teacher candidates, Minor, Onwuegbuzie, Witcher, and James (2002) found that 28.4% believed the role of the teacher is to transmit information to students while only 12.7% held the progressive belief that students should be actively engaged in solving learner-generated problems. Experiences with number strings could support progressive beliefs and higher levels of math talk because an important element of the number string routine—applying reasoning (Chapin et al., 2009)—supports a shift in responsibility from teacher to student by encouraging students to help their classmates under-

stand the mathematics, even agreeing or disagreeing with other students' reasoning when necessary.

Research Questions

Based on the elements that are part of the number string protocol, we hypothesized that utilizing them could help teacher candidates facilitate higher levels of math talk. However, even though the number string protocol includes steps for addressing different levels of questions, focusing on explanations, and letting students take a more central role, novice teacher candidates might struggle to enact all parts of the protocol. Therefore, it is important to know what parts of the protocol teacher candidates do enact and to what extent they can support higher levels of math talk, leading us to pose the following research questions:

1. To what extent do elementary teacher candidates enact the elements of the number string protocol in an initial lesson?

2. What levels of math talk are they able to facilitate across the four categories?

3. What are their successes and struggles in enacting number strings meaningfully in terms of the levels of a math-talk learning community?

Methods

Participants and Setting

The participants were 24 master's degree students taking a mathematics methods course as part of a 1-year, full-time elementary education teacher training program in California. The teacher candidates assisted in placement classrooms during the morning and took classes at the university in the afternoon. In their placement classrooms, the teacher candidates worked with individual students or small groups around mathematics and taught several mathematics lessons to the whole class. The mathematics methods course met once per week for 3 hours. We selected four teacher candidates (Zoe, Lana, Becca, and Stephanie—all names are pseudonyms) for our focus because they all worked in third grade, English-speaking classrooms. According to Zoe and Stephanie, students in their classrooms often did mental math problems where they could write their work on white boards and compare answers, but only students in Zoe's classroom regularly *shared* their strategies. Students in Lana and Becca's classrooms did not regularly participate in mental math activities as a class. Three of the four teacher candidates designed and used the same number string. Eliminating variation helped us focus on the implementation of the number string in the teacher candidates' placement classrooms as well as the math-talk levels that occurred during the number strings.

Modeling Number Strings in Methods Class

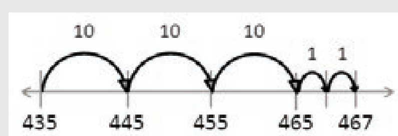
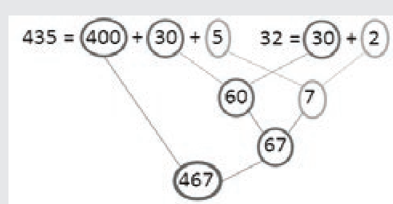
A major goal of this methods class was to prepare the teacher candidates for leading a student-centered series of mathematics lessons with a focus on eliciting student thinking, and we felt the number string routine had the potential to help teacher candidates elicit and build on student thinking in an accessible way. At the beginning of each class session throughout January and February, the instructors conducted a number string (see **Appendix A** for examples) with the teacher candidates to familiarize them with the protocol, as described by Kazemi, Franke, and Lampert (2009). The number string problems were selected to be hard enough to give the teacher candidates an authentic number string experience. In Table 1, we present excerpts from two example number strings conducted with the teacher candidates that highlight how the instructors incorporated elements of the number string protocol.

As seen in the excerpt from String B (see Table 1), the instructors introduced and talked about open number lines and branching notation as key representations to use to illustrate students' strategies. In String A and other multiplicative examples, they highlighted the use of arrays.

Data Sources

After participating in six number strings, during which the instructors explicitly discussed the components of the routine, the teacher candidates designed and eventually taught a number string in their practicum placement. For the assignment, they had to (1) identify the big idea for their string, (2) develop or find a number string, (3) anticipate students' strategies (both correct and emerging), (4) identify possible representations they could use to illustrate students' strategies, and (5) list possible questions they could use to facilitate the discussion around the problems. The teacher candidates had read sections of Sherry Parrish's book, *Number Talks: Helping Children Build Mental Math and Computation Strategies (Grades K–5)*; however, most candidates designed their strings based on the examples from class and the level of their students in their practicum placements. They worked together in grade-level teams to design strings appropriate for their students.

Table 1
Class Number String Descriptions (I = instructor, TC = teacher candidate)

Excerpts from Demos		Instructor's role
String A: *Instructor writes 15 x 18 on the board.*		Writes problem horizontally.
I:	Think about how you might mentally find the product.	Invites participation.
TC:	First, I knew 10 x 18 is 180. Then I knew 5 x 18 was half that, so 90. Then I added 180 + 90 to get 270.	*Instructor writes equations on the board.*
I:	Is there another way to think about the problem?	Elicits multiple strategies.
TC:	I knew that 15 x 20 is 300, and that is two too many fifteens, so I subtracted 30 from 300 to get 270.	*Instructor writes equations on the board.*
I:	It's wonderful how many different ways there are of thinking about this problem.	Places an emphasis on the celebration of diverse thinking.
TC:	I cut 18 in half to get 9 and doubled 15 to get 30. Then I multiplied 9 x 30 to get 270.	*Instructor writes equations on the board.*
I:	That is an interesting approach. Will you please tell me more about why you did that?	Demonstrates how to handle unclear strategies by probing.
TC:	[*Repeats statement without adding anything.*]	
I:	Who can use their own words to explain the strategy?	Invites others to restate.
Two more candidates explain the doubling and halving strategy using different words.		
I:	Thank you to everyone for helping us all understand this strategy. At first I wasn't sure about the strategy, but our discussion has helped clarify things more for me.	Emphasizes class participation.
TC:	[*Pretending to be a student*] I got 140 because 10 x 10 is 100 and 8 x 5 is 40.	
I:	Thank you for sharing. [*The instructor drew an array to illustrate the components of the product and show the missing parts before writing the next problem: 18 x 27.*]	Accepts incorrect answers and uses representations to illustrate strategies.
String B: *Instructor writes 435 + 32 on the board.*		Writes problem horizontally.
TC:	I made three jumps of ten and then two jumps of one.	Shows the jumping strategy using an empty number line. 
TC:	First, I added 30 + 30, then 5 + 2. I took that 67 and added it to 400 and got 467.	Shows the recomposing strategy using branching lines. 

Once candidates developed their number strings, the instructors provided feedback on them. The most common feedback was to begin the string with an easier problem—selection of problems that were too difficult came perhaps from the fact that the number strings done in the methods class were designed for the teacher candidates. Other feedback included ideas on connecting possible student strategies as well as using visual representations to capture possible student thinking. In their written reflections, teacher candidates stated that this feedback, especially suggestions for the first problem of the string, was very helpful. Each candidate then taught the number string in her placement classroom and submitted a video of her enacted number string and a reflection that described (1) the actual strategies students used during the number string, (2) any changes they made to the problems as the string progressed, (3) descriptions of their experiences using representations to highlight student thinking, and (4) a discussion of changes they would suggest or insights they had that would be helpful to others using their number string.

Data Analysis

The second author transcribed each number string video. After dividing each lesson video transcript into chunks consisting of a problem introduction or a strategy request plus a response, we used the descriptions of the math-talk levels (see Hufferd-Ackles et al., 2004) to code the level of questioning, explaining, source, and responsibility evident in each exchange. It should be noted that the teacher candidates were not familiar with the math-talk levels; these were used for research purposes only. As we coded the first participant, we found that some responses were not solidly at one level but fell in between two sequential levels. After discussion, we marked these as transitioning between the two levels to capture these nuances (see Table B1 in **Appendix B** for descriptions of the levels). We individually coded the remaining participants and had 77%–81% agreement across the four categories. Next, we discussed the discrepancies until we reached consensus (see Table 2 for an example of our coding).

The teacher candidates' planning notes provided insight into the big idea for their strings, and their reflections were coded for references to the four math-talk categories as well as the elements of the number string protocol, which enabled us to triangulate the results of our coding with the teacher candidates' thoughts about their number strings.

Results

Enacting Number Strings

Lana, Becca, and Stephanie used the problems 4 + 7 and 44 + 7 to encourage students to break apart numbers, regroup them, and notice similarities in the ones places.[2] Zoe, at her cooperating teacher's request, taught a different string and emphasized a common difference strategy using 25 – 3, 25 – 13, 26 – 14, and 28 – 16. All four candidates successfully focused on students' strategies rather than just their answers and were able to use a variety of representations to make the strategies visible to the students. Although the teacher candidates facilitated some conversations around making connections among strategies, highlighting the big idea of the number string was more challenging. Table 3 provides a summary of number string elements that the candidates enacted.

Math-Talk Levels Across Categories

Overall, the number string routine provided a helpful structure for all four teacher candidates to progress beyond a level 0 teacher-focused community toward a level 1 student-focused math-talk community. Their average percentage of math talk for each level of the four categories is shown in Table 4. The prevalence of level 1 codes for *questioning* reflects the candidates' emphasis on asking students about how they solved the problems; if students had asked each other more questions, the level for questioning would be higher. Similarly, the high percentage of *explaining* at level 1 occurred because the candidates frequently had to ask several follow-up questions in order to obtain a complete explanation; students did not offer full explanations on their own. The most common level of coding for *source* (1–2) reflects the candidates' efforts to build on students' thinking by representing their strategies in various ways. The most common level of coding for *responsibility* (0–1) occurred because although students were listening to each other's thinking, they were not yet initiating or leading the conversations.

Next we give examples of the candidates' use of the math-talk components and discuss practices that supported or hindered their efforts to move toward higher levels of math talk.

2 None of the three teacher candidates got to their other three planned problems: 49 + 7, 49 + 17, and 56 + 18.

Table 2
Example Codes for a Transcript Excerpt

Transcript for lower levels of math talk	Codes based on Hufferd-Ackles et al. (2004)
[*One student shares a strategy for solving 7 + 4 prior to this excerpt.*] *Stephanie:* Did someone do it differently? *Student:* I looked at 7 plus 4 and you could break 4 into 3 plus 1. *Stephanie:* Okay, so you knew that 3 plus 1 is 4. *Student:* . . . and then I did 7 plus 3 and then added the 1. *Stephanie:* So you knew that 7 plus 3 is? *Student:* 10. *Stephanie:* So you knew that was 10. And then you added the 1 to get 11? *Student:* Yes. *Stephanie:* Oh, this is exciting—we're solving it in interesting ways.	**Questioning:** Level 1 (Teacher allows student to explain and is interested in the response.) **Explaining:** Level 1–2 (Student gives a more complete explanation with some details missing initially.) **Source:** Level 1–2 (Teacher expands on student's explanation.) **Responsibility:** Level 0–1 (Teacher elicits multiple strategies, but other than that, students are not involved in talking about strategies.)
Transcript for higher levels of math talk	**Codes based on Hufferd-Ackles et al. (2004)**
[*After discussing how 28 – 16 and 26 – 14 are related*] *Zoe:* Okay, last comment, Dorothy? *Student:* I just wanted to say to Alicia that her answer of 22 could not be possible because you have 28 and you're minusing 16. There is no possible way to get 22. *Zoe:* We're just off by a one [ten]. We're all really close today.	**Questioning:** Level 1 (Teacher allows student to explain and is interested in the response.) **Explaining:** Level 2–3 (Student gives an unprompted statement that stakes a position, but teacher does not advance it.) **Source:** Level 3 (Teacher allows students to interrupt and provide their own explanations or interpretations.) **Responsibility:** Level 3 (Students agree or disagree with each other without prompting and work together to establish the answers to problems.)

Questioning

All four candidates asked level 1 questions, such as "How did you get your answer?" Notably, all candidates let students' explanations dictate correctness. Also, because they asked, "Who solved the problem a different way?" students began to take responsibility for listening to their peers so as not to repeat strategies.

Low press. All candidates used low-press questions to encourage clarification: "So you started with seven and put four in a group. What did that look like in your head?" (Becca) "You counted by ones. What did you start with?" (Stephanie)

High press. All four candidates used at least one high-press question during their number string. When one student noticed that 12 was the answer to all of the problems, Zoe asked the students to determine the reason for the pattern. However, most high-press questions focused less on patterns and more on clarifying the mathematics in explanations, especially around splitting or decomposing and recomposing numbers. For 7 + 4, one student said that she added 7 + 3 to get 10 and then added 1 more to get 11. Becca asked, "So how did you know to add one more after you added three?"

These high-press questions drew attention to the decompositions and set an expectation for the students to justify their strategies, and the low-press questions encouraged students to explain their strategies in more detail. By pursuing students' strategies in more depth, these teacher candidates communicated to the class that their ideas formed the basis of their learning. Further, Zoe was

Table 3
Number String Elements Enacted by the Teacher Candidates

	Zoe	Lana	Becca	Stephanie
Video length (in minutes)	25	22	18	28
Number of coded chunks	24	12	10	9
Number of strategies elicited	13	10	8	10
Probed student thinking	12	35	24	29
Addressed mistakes	4	4	0	0
Highlighted friendly numbers	1	3	1	2
Made strategy connections	1	1	1	3
Use of representations				
Number lines	3	0*	2	8
Recomposing lines	5	0	0	0
Pictures (e.g., dots)	0	4	1	0
Written equations only	4	5	5	2

* In her reflection, Lana indicated that she used number lines, although they were not visible in the video.

Table 4
Average Percentage of Math Talk Across the Four Teachers

Math Talk Levels	Categories			
	Questioning	Explaining	Source	Responsibility
0	11%	13%	11%	0%
0–1	19%	8%	6%	68%
1	64%	46%	9%	9%
1–2	6%	13%	57%	13%
2	0%	17%	15%	8%
2–3	0%	2%	0%	0%
3	0%	0%	2%	2%

successful in transitioning toward a level 2 as she asked, "Should we ask our friends if they know anything? Do you want to call on a friend?" to encourage peers to ask each other questions.

Challenges. Although low-press questions came naturally to the candidates, they sometimes asked leading questions, limiting students' responses. For example, one student said that she saw 4 + 7 as seven dots and four more. Lana asked her if she counted (assuming this was her method); a more generative question would have been, "What did you do next?" Lana then referred to an illustration of dots on the board, demonstrating, "I'm going to start [counting] from here: 1, 2, 3 . . . ? Or did you say 7, 8, 9, 10, 11?" Asking the student, "How did you count?" would have given the student an opportunity to explain more deeply.

Explaining Mathematical Thinking

Across the candidates' number strings, 46% of their explanations of mathematical thinking were characteristic of level 1 math talk, with 30% of their explanations transitioning to or reaching level 2. As the candidates continued to probe, students' explanations became more detailed with less prompting. Lana explained, "When students explained their thinking as 'just knowing' the answer, I pushed them further by asking them to picture what they saw in their minds and really think about the process they went through to get the answer. . . . When this [more detailed explanation] happened, other students

started to recognize their own processes and the discussion became richer."

Revoicing. The teacher candidates played a strong role in eliciting student thinking and then revoicing students' explanations to encourage more complete answers. The candidates' use of revoicing fell into several categories. In some cases, they repeated students' statements verbatim; this served to make sure all students heard their peers' ideas. Other times, candidates revoiced and named a student's strategy or used more precise mathematical language, which gave legitimacy to the students' explanations:

(Solving 25 – 3)

> *Student:* I pretended that the 2 wasn't there, and I knew that 5 minus 3 equals 2 and then I just added the 2 and it equaled 22.
>
> *Zoe:* So you kinda used our *canceling* out strategy up there. So you pretended that it was 5 minus 3 and pretended that [2] wasn't there and then you put it back.

Zoe acknowledged that the student was using the canceling strategy that the class had talked about previously for non-regrouping problems, making connections to the class's prior knowledge.

Challenges. Even though all four candidates probed their students' explanations during the number strings, they often over-revoiced, which took away the opportunity for students to do the bulk of the explaining. This shifted the focus away from the students and back to the teacher and moved the discourse community back toward level 0 math talk, where the teacher is considered the source of mathematical ideas and holds the responsibility for student learning. Although the candidates' methods instructors had talked to them about listening to the students, revoicing, asking questions, and not telling too much, the candidates had not discussed what is considered too much prompting or revoicing.

The candidates' over-revoicing may have resulted from the difficulty they thought students experienced in response to their probing. For instance, in her reflection, Zoe wrote, "I asked a student to explain how she knew what to do with the remaining numbers, but she seemed a bit confused by the question, so I moved on." Becca said, "I'm not sure if the students' difficulties explaining their answers result from a lack of experience doing mental math with various strategies, or if they were confused about their own thought processes." Despite students' struggles explaining their mathematical thinking, however, the teacher candidates seemed optimistic about the effect of consistently doing number strings. Becca hypothesized, "I expect that [students] would develop the needed mathematical vocabulary with more frequent math string activities."

Source of Mathematical Ideas

A large percentage (57%) of the segments in the number strings were coded Level 1–2 in the "source" category, making it the most successful component. During their number strings, the teacher candidates elicited multiple strategies. More importantly, the candidates allowed students to use their own strategies and make errors and gave students opportunities to identify their miscalculations rather than telling them the correct answers. This shifted the source of mathematical ideas away from the teacher candidate and onto the students.

Representing student thinking. The candidates legitimized students' strategies by representing them visually. Overall, the teacher candidates used empty number lines, branching/recomposing lines, pictures, and/or equations to illustrate students' strategies, especially when the students' strategies were easy to understand. Zoe used empty number lines when students talked about counting back or making jumps and used recomposing or branching lines whenever students talked about breaking apart a number, even if the student only broke apart one number and then made jumps. When solving 26 – 14, one student suggested subtracting 4 from the 26 and 4 from the 14 and then subtracting the two resulting answers. Zoe wrote the horizontal equations 26 – 4 = 22, 14 – 4 = 10, and 22 – 10 = 12 vertically stacked on the board.

Connections among strategies. After students shared their strategies, the candidates found ways to help students make connections among strategies. For example, when one student explained an important strategy for solving 28 – 16, Zoe capitalized on it to make connections to 26 – 14 and the big idea of the number string:

> *Student 1:* I put 28, and right next to 28 I put a 16, and then minused 6 [from 28] and it equaled 22, and then 16 minus 6 equals 10, and then 22 minus 10 is 12.
>
> *Zoe:* So [Student 1] noticed that we can do the same thing to both numbers. So 28 minus 6 and then 16 minus 6 to get 10, a friendly number. And then he subtracts the two numbers. What did we do to get from 26 to 28?

Student 2: We just added 2.

Zoe: What did we do to get from 14 to 16?

Student 2: We also added 2.

Zoe: That's kind of what [Student 1] did, only in reverse.

Lana also helped students make connections among the ways to make 10 when adding 4 + 7, depending on if the student started with 7 or 4. According to her reflection, she felt that "this was a great opportunity to compare the two equations and show how making tens in regrouping seemed to be a trend." Later, for 44 + 7, when a student made a multiple of 10 by adding 44 + 4 = 48 and then 48 + 2 = 50, she told the class, "Are you noticing a pattern here? People are wanting to make tens."

In other cases, students made connections to each other's strategies. While working on the problem 4 + 7 and after one student had found 10 as a friendly number (7 + 3) and then added 1 to get 11, Stephanie asked her students:

Stephanie: Did anyone do it differently?

Student: I did it like [the previous student], but I did 6 plus 4 is 10. And then added on.

The student was able to connect his thinking to another student's regarding ways of finding 10 as a friendly number, which further solidified that the students were the ones presenting the mathematical ideas. In her reflection, Stephanie said, "Thirty minutes and two problems into this number string—*after* I had informed the class that it was time for recess—many students were still sharing their ideas." Her statement shows that not only were students able to be the source of mathematical thinking during the number string, but they were also enthusiastic about doing so.

Challenges. Although the teacher candidates helped students *make* some connections among strategies, they had a difficult time *showing* these connections with their representations. For example, writing the equations 26 – 4 = 22, 14 – 4 = 10, and 22 – 10 = 12 vertically stacked on the board did not help Zoe show why 26 – 14 and 22 – 10 have a common difference. If she had written each problem vertically side by side (Bofferding, 2012) or on a number line, she could have highlighted that this student was keeping a common difference (see Figure 1).

Showing connections requires that candidates know multiple ways to represent each strategy and can decide which representation will be most effective for a particular situation, which may require additional experience.

Responsibility for Learning

Overall, 68% of the exchanges were coded at a Level 0–1 for the responsibility component. Zoe had a nice moment, however, near the end of her number string that met the criteria for a Level 3 exchange. She put the problem 28 – 16 up on the board, and students initially presented three different answers (12, 22, and 11). Four students shared their strategies, and the consensus seemed to be that the answer was 12. Then one student raised her hand, and Zoe called on her:

Zoe: Okay, last comment. Dorothy?

Student: I just wanted to say to Alicia that her answer of 22 could not be possible because you have 28 and you're minusing 16—there is no possible way to get 22.

While this could, on the surface, be seen as an affront to Alicia, Dorothy's tone of voice indicated that she was trying to be helpful to her classmate. This depicts a subtle shift in responsibility away from the teacher and onto the

$$\begin{array}{l} 26 - 14 = 12 \\ \underline{-4} \quad \underline{-4} \\ 22 - 10 = 12 \end{array}$$

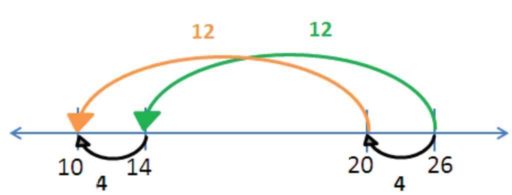


Figure 1. **Two ways to highlight how to maintain a common difference.**

students. They were beginning to care about each other's learning, and they wanted to help each other understand. In her reflection, Stephanie said, "I really enjoyed hearing some of the students—particularly those who normally like to work on their own—respond to their classmates' strategies and compare their solutions to the work on the board." While there is still much room to grow in this area, three of the candidates specifically encouraged collaboration during their number string. Zoe asked, "Do you want to call on a friend?" Becca asked, "Does anyone know where [Student] was going with this?" Stephanie asked, "Can anyone explain [Student's] way?" These types of probes encouraged students to take responsibility for understanding each other's mathematical thinking.

Challenges. This component of a math-talk community was the most challenging for the teacher candidates during their number strings. The prevalence of level 0–1 codes indicates that teacher candidates held onto the notion that they ultimately were responsible for managing and facilitating the mathematical discussion. While there were moments when students directed their comments to each other or referenced a peer's strategy, the bulk of the exchanges were between teacher and student, with the teacher directing the flow of the conversation.

Discussion

Number String Elements and Math-Talk Levels

Overall, the results provide an encouraging account of the power of number strings in helping teacher candidates use more student-centered levels of math talk. Because the number string routine requires that teachers ask students how they solved the problems and record their strategies, teachers naturally asked prompting questions to make sure they understood how to represent students' strategies accurately. This practice helped them be better prepared to deal with unanticipated responses. Rather than ignoring responses (Inoue & Buczynski, 2011), the candidates tried to record students' thinking on the board and asked clarifying questions. Although the candidates did not always feel they truly understood the students' methods, they actively tried to and, by recording the methods, provided an opportunity for other students to try to clarify the methods as well.

Later in their number string conversations, students then began to reference each other's work or make explicit comparisons between their strategies. Rather than the teacher determining correctness (Muis, 2004), the students had a more active role. Another math-talk strategy that likely contributed to students' participation was candidates' specifically asking students to make sense of what their classmates said or to add on to each other's statements. Both experiences provided students with reference points on which to build. All this is evidence that even within one number string, all four candidates were able to move the class toward higher levels of math talk.

The teacher candidates used revoicing naturally (and liberally) throughout their lessons; however, they could increase their use of restating, having other students summarize each other's strategies. This would provide students with additional opportunities to actively engage with their peers' thinking. Although encouraging students to build on each other's thinking did help students start to take responsibility for their learning, the *responsibility* category was consistently coded the lowest in terms of the math-talk levels. This is not surprising because this category is the most dependent on establishing a classroom culture based on student-centered practice. Such extended collaboration is not likely to surface in one initial lesson. However, teacher candidates can encourage collaboration through partner talk and a more targeted focus on helping students make connections among strategies. Three types of comparisons emerged in these cases: the teacher relating students' strategies to the big idea of the string, the teacher comparing two or more students' strategies, and students making comparisons to peers' strategies. Beyond these, though, teacher candidates need to capitalize on the comparisons, especially in terms of having the class analyze the efficiency and utility of their strategies.

Teacher candidates can also encourage students to take responsibility by providing more opportunities for students to agree and disagree by addressing students' mistakes. Candidates often had students explain their strategies in the order that they volunteered their answers. If the second student they called on had an incorrect answer, it was often not probed because the correct answer was already discussed; therefore, two candidates did not address any mistakes. Candidates could have students who presented incorrect answers describe their solution processes, even though they were incorrect, to give the rest of the class a chance to disagree; alternatively, they could point out that the class got several possible answers and ask the students to agree or disagree with one of the answers and explain why. Having students think critically about the answers and strategies would help them transition to taking more responsibility for the conversation.

Common Core Mathematical Practices

In terms of the Common Core State Standards for Mathematical Practice (National Governors Association Center

for Best Practices & Council of Chief State School Officers, 2010), the results of this study provide some insights into ways to help students construct viable arguments and critique the reasoning of others (practice #3). One key practice that all teacher candidates used was recording students' strategies on the board using equations, number lines, and branching notation, which teaches students how to use appropriate tools strategically (practice #5) and how to model with mathematics (practice #4).

Using the Number String Protocol in Methods Classes

The ease with which the teacher candidates enacted the number string protocol suggests that this type of lesson is a powerful first step in learning to facilitate productive mathematical discussions with students. Well-thought-out protocols, such as the number string protocol, can provide a structure that helps beginning teachers incorporate more student-centered practices into their teaching.

However, the number string structure does not guarantee that all of the discourse will be rich. Lana in particular struggled to move away from the role of teacher as the source of mathematical ideas. Although she successfully moved away from teaching as telling, about half of her questions led students to answer in particular ways (Nicol, 1999). A discussion in methods courses about leading versus open questions could bring these challenges to the surface, especially when combined with rehearsals (Kazemi, Franke, & Lampert, 2009) focused on the use of questioning.

In contrast to prior reports about teacher candidates having difficulty understanding the mathematics content (e.g., Nicol, 1999), the teacher candidates in our study expressed frustration with understanding students' strategies, particularly as it related to representing their strategies. Providing opportunities for teacher candidates to rehearse and represent strategies "on the spot" in methods courses could give them experience representing while others are explaining yet also give them a space where they can stop and ask for suggestions and the class can problem-solve effective representations for various strategies.

Although all candidates were able to elicit multiple strategies from students, the end goal of the number strings was not always clear. Only Zoe specifically moved students toward thinking about the constant difference pattern among problems in their string. One suggestion to improve candidates' and students' attention to the relations among the problems and strategies would be to identify specific questions instructors could incorporate as necessary elements into the number string protocol to promote the higher levels of talk. For example, (1) after students share their strategies for a problem, ask them to describe how their strategy is similar to or different from another student's, (2) after solving a few problems, have students identify how two of the previous problems are similar and different, and (3) at the end of the string (if no one has noticed the pattern in the problems), ask, "How could you use the first problem to help you solve the second?" (Wickett, 2003).

Limitations and Future Work

While this study examines the math-talk levels that occur during four teacher candidates' first number string enactment with their students, there is no comparison data that allows us to look at the math-talk level of these teachers during other kinds of math lessons. We also do not know how their talk compares to average new teachers. A next step would be to not only videotape and code teacher candidates' number string lessons, but also other math lessons they teach to see if there are differences and/or spillovers from the number string protocol. Additionally, it would be interesting to compare the math talk in number strings and math lessons to discussion/talk in other subjects. Are there cross-curricular discussion skills that can help teachers engage their students in meaningful talk across subjects?

Given the success the teacher candidates had implementing number strings, we should continue to explore issues surrounding their use. Future work should also explore when (if at all) leading questions are beneficial, what factors influence the types of questions teachers ask, and explore instructional activities, such as rehearsals, that can help support teachers to move toward more open questions that will enable them to elevate the math talk in their classrooms. Additionally, it would be meaningful to follow teacher candidates into the first 2–3 years of their career to see to what extent they are able to maintain or build on what they have learned from the number string protocol.

References

Boaler, J., & Brodie, K. (2004). The importance, nature, and impact of teacher questions. In D. E. McDougall & J. A. Ross (Eds.), *Proceedings of the 26th Conference of the Psychology of Mathematics Education North America* (Vol. 2, pp. 773–782). Toronto, Canada: OISE/UT.

Bobis, J. (2007). The empty number line: A useful tool or just another procedure? *Teaching Children Mathematics*, *13*(8), 410–413.

Bofferding, L. (2012, October–November). The continuum of pre-service teachers' MKT of math strings. In L. R. Van Zoest, J. Lo, & J. L. Kratky (Eds.), *Proceedings of the 34th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 697–700). Kalamazoo, MI: Western Michigan University.

Borg, M. (2004). The importance of observation. *English Language Teaching Journal*, *58*(3), 274–276.

Brookhart, S. M., & Freeman, D. J. (1992). Characteristics of entering teacher candidates. *Review of Educational Research*, *62*(1), 37–60. doi:10.3102/00346543062001037

Chapin, S. H., O'Connor, C., & Anderson, N. C. (2009). *Classroom discussions: Using math talk to help students learn* (2nd ed.). Sausalito, CA: Math Solutions.

DiBrienza, J., & Shevell, G. (1998). Number strings: Developing computational efficiency in a constructivist classroom. *The Constructivist*, *13*(2), 21–25.

Fosnot, C. T., & Dolk, M. (2001). *Young mathematicians at work: Constructing multiplication and division.* Portsmouth, NH: Heinemann.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, *111*(9), 2055–2100.

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal*, *30*(2), 393–425.

Hufferd-Ackles, K., Fuson, K. C., & Sherin, M. G. (2004). Describing levels and components of a math-talk learning community. *Journal for Research in Mathematics Education*, *35*(2), 81–116.

Inoue, N., & Buczynski, S. (2011). You asked open-ended questions, now what? Understanding the nature of stumbling blocks in teaching inquiry lessons. *The Mathematics Educator*, *20*(2), 10–23.

Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, *41*(2), 169–202. Retrieved from http://www.jstor.org/stable/20720130

Kazemi, E., Franke, M., & Lampert, M. (2009). Developing pedagogies in teacher education to support novice teachers' ability to enact ambitious instruction. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 12–30). Palmerston North, New Zealand: MERGA.

Kazemi, E., & Stipek, D. (2001). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *The Elementary School Journal*, *102*(1), 59–80.

Kemmerle, M. (2013, November). Promoting student questions: A high school mathematics case study. In M. V. Martinez & A. C. Superfine (Eds.), *Proceedings of the 35th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1004–1011). Chicago, IL: University of Illinois at Chicago.

Klein, A. S., Beishuizen, M., & Treffers, A. (1998). The empty number line in Dutch second grades: Realistic versus gradual program design. *Journal for Research in Mathematics Education*, *29*(4), 443–464.

Lampert, M., Beasley, H., Ghousseini, H., Kazemi, E., & Franke, M. (2010). Using designed instructional activities to enable novices to manage ambitious mathematics teaching. In M.K. Stein & L. Kucan (Eds.), *Instructional explanations in the disciplines* (pp. 129–141). New York, NY: Springer.

Minor, L. C., Onwuegbuzie, A. J., Witcher, A. E., & James, T. L. (2002). Preservice teachers' educational beliefs and their perceptions of characteristics of effective teachers. *The Journal of Educational Research*, *96*(2), 116–127. doi: 10.1080/00220670209598798

Muis, K. R. (2004). Personal epistemology and mathematics: A critical review and synthesis of research. *Review of Educational Research*, *74*(3), 317–377. Retrieved from http://www.jstor.org/stable/3516027

Murata, A., Siker, J., Kang, B., Baldinger, E., Kim, H., Scott, M., & Lanouette, K. (under review). Classroom group learning and individual student learning through math talk: The case of two first grade classrooms. *American Educational Research Journal.*

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics.* Washington, DC: Author.

National Research Council. (2001). *Adding it up: Helping children learn mathematics.* J. Kilpatrick, J. Swafford, & B. Findell (Eds.) Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

Nicol, C. (1999). Learning to teach mathematics: Questioning, listening, and responding. *Educational Studies in Mathematics*, *37*(1), 45–66.

O'Connor, C., & Michaels, S. (2007). When is dialogue "dialogic"? [Commentary]. *Human Development*, *50*, 275–285. doi:10.1159/000106415

O'Loughlin, T. A. (2007). Using research to develop computational fluency in young mathematicians. *Teaching Children Mathematics*, *14*(3), 132–138. Retrieved from http://www.jstor.org/stable/41199082

Parrish, S. D. (2011). Number talks build numerical reasoning. *Teaching Children Mathematics*, *18*(3), 198–206.

Reinhart, S. C. (2000). Never say anything a kid can say! *Mathematics Teaching in the Middle School*, *5*(8), 478–483.

Smith, J. P. (1996). Efficacy and teaching mathematics by telling: A challenge for reform. *Journal for Research in Mathematics Education*, *27*(4), 387–402. Retrieved from http://www.jstor.org/stable/749874

Stein, M. K., Engle, R., Smith, M., & Hughes, E. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, *10*(4), 313–340.

Wickett, M. S. (2003). Discussion as a vehicle for demonstrating computational fluency in multiplication. *Teaching Children Mathematics*, *9*(6), 318–321.

AUTHORS

Laura Bofferding, Purdue University, 100 N. University Street, Beering Hall #4132, West Lafayette, IN 47907; LBofferding@purdue.edu

Melissa Kemmerle, Stanford University, 485 Lasuen Mall, Stanford, CA 94305; Kemmerle@stanford.edu

Appendix A: Example Number Strings Used in the Methods Course With the Teacher Candidates

1. Find Friendly Numbers (Decomposing Branching Notation)

   35 + 36

   75 + 36

   37 + 59

   175 + 46

2. Making Jumps to Helpful Numbers (Number Lines)

   1999 – 4

   2002 – 4

   2002 – 14

   1898 – 14

3. Compensating (x 20 and Subtract Groups); Using Place Value (Arrays)

   15 x 20 (*not given, but would help prompt the targeted strategy)

   15 x 18

   24 x 18

(**Return to page 102**)

Appendix B: Example Explanations of Codes

Table B1

Explanations of Codes Used in Data Analysis of Teacher Candidates' Lessons

Level	Questioning	% Math talk
0	Teacher asks question for sole purpose of getting answer (no inquiry into students' methods).	11%
0–1	Teacher asks a question that indicates an interest in students' explanations, but follows up by asking (mostly) leading questions instead of allowing students to do most of the explaining.	19%
1	Teacher asks questions about how students got their answers (and why they did what they did). Teacher allows students to explain.	**64%**
1–2	Teacher encourages students to think about questions they have about their peers' explanations.	6%
2	Teacher prompts students to ask each other questions.	0%
2–3	Student initiates questions but does not follow up on the responses.	0%
3	Student initiates questions about classmates' work and probes the responses.	0%
	Explaining Mathematical Thinking	
0	Student provides an answer to a math problem but offers no explanation of his or her thinking.	13%
0–1	Student names the strategy he or she used but provides little detail on what this entailed.	8%
1	Student gives more explanation about the details of the strategy, but teacher does a lot of helping (asks leading questions or adds too much detail when revoicing what the student said).	**46%**
1–2	Student gives a more complete explanation with some imprecise language.	13%
2	Student gives a full explanation; teacher doesn't have to clarify much.	17%
2–3	Student gives an unprompted statement that stakes a position, but teacher does not advance it.	2%
3	Student provides justification as part of an explanation and teacher encourages deeper thinking about strategies.	0%
	Source of Mathematical Ideas	
0	Student only gives answer and doesn't expand upon it.	11%
0–1	Student gives barely more than an answer, but it is evident that he or she is beginning to think about the strategy (instead of the answer only).	6%
1	Student gives an explanation with little response or building on by the teacher.	9%
1–2	Teacher expands on student's explanation by illustrating it with a representation, naming a strategy, or helping the students start to compare; may attempt to address errors but in a vague way. Student is more confident.	**57%**
2	Students are beginning to take charge and explain their thinking to each other. Teacher takes advantage of students' errors as opportunities for learning.	15%
2–3	Teacher begins to use students' ideas for building new understanding.	0%
3	Teacher allows students to interrupt and provide their own explanations or interpretations.	2%

Responsibility for Learning		
0	Teacher affirms students' answers as correct or incorrect (not observed).	0%
0–1	Teacher elicits multiple answers, and there is some sense that students need to listen (because they will have to say whether they have a different or similar answer). Students listen to each other.	**68%**
1	Student references or repeats what another student said but provides little further explanation.	9%
1–2	There is some sense of comparison or analysis between strategies. A student might reference a classmate's strategy and say something such as "I did another jump."	13%
2	Students work together to understand and get the answer at the teacher's prompting.	8%
2–3	Students agree or disagree with each other but need prompting, especially to decide on answers.	0%
3	Students agree or disagree with each other without prompting and work together to establish the answers to problems.	2%

(**Return to page 104**)

Reshaping Teachers' Mathematical Perceptions: Analysis of a Professional Development Task

Gwyneth Hughes
Boise State University

Jonathan Brendefur
Boise State University

Michele Carney
Boise State University

As the focus of mathematics education moves from memorization toward reasoning and problem solving, professional development for in-service teachers must model these activities while simultaneously increasing participants' mathematical knowledge. We examine a representative task from a mathematics professional development course that uses rational number operation as an opportunity for problem solving and modeling. Transcripts exemplify the growth teachers make in deeply understanding the content—division of fractions—while engaging in guided reinvention and classroom discourse. We propose 4 interconnected qualities of this task that allow participants to engage in and reflect on the process of guided reinvention: (1) authentic context with multiple solution methods, including visual; (2) cognitive dissonance; (3) deep engagement; and (4) impact on mathematical knowledge for teaching.

Keywords: Content knowledge; Fraction division; Mathematical knowledge for teaching; Professional development

Since the publication of *A Nation at Risk* in 1983 (Gardner, 1983), multiple educational institutions and researchers within the United States have called for increased rigor in mathematics instruction and a shift from a strict focus on procedural fluency to incorporation of problem solving and critical thinking (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010; National Council of Teachers of Mathematics, 1989; National Council of Teachers of Mathematics, 2000; Stigler & Hiebert, 1999). Several authors have suggested that tasks that progressively formalize number operations can be used as sense-making activities in and of themselves (Fosnot & Dolk, 2002; Gravemeijer & van Galen, 2003; Hiebert, 1997). The outcome for students of such lessons would be a combination of procedural fluency and conceptual understanding specific to the topic, as well as general experience with the practices of mathematics, including conjecture, reasoning, justification, and modeling.

But such classroom activity can be difficult for teachers who themselves experienced mathematics instruction that focused on memorization of formal algorithms (Ball & McDiarmid, 1989; Schoenfeld, 1988). Teaching mathematics rather than computation requires several different types of knowledge, collectively referred to as mathematical knowledge for teaching (Ball, Thames, & Phelps, 2008; Hill, Rowan, & Ball, 2005). Teachers must understand the conceptual basis for number operations *and* be able to construct and facilitate problem-solving situations for students around those operations. Building upon student thinking also requires a fundamental shift in teachers' beliefs about the nature of mathematics itself, from that of mathematics as a set of rules and algorithms to mathematics as a product of quantitative reasoning (Ernest, 1989; Schoenfeld, 1988). The widespread adoption of the Common Core State Standards (CCSS) and the accompanying assessments has, to some degree, forced this shift because students must demonstrate content knowledge and engage in authentic practice of mathematics.

Professional development for in-service teachers has been widely seen as one method for increasing teachers' mathematical knowledge for teaching (Borko, 2004; Hill & Ball, 2004). Here we describe a single activity from a mandated professional development course—the Mathematical Thinking for Instruction (MTI) course. Within this single problem, we identify attributes of the task (and thus the course) that allow teachers to experience the process of guided reinvention as though they were students. The purpose of this paper is to (1) describe the MTI professional development course in order to situate the task; (2) describe the task and its facilitation, including class transcripts and participant-generated figures; and (3) examine participants' course evaluation reflections in order to identify attributes of the intervention that make it effective in modeling guided reinvention.

The Mathematical Thinking for Instruction (MTI) Course

In 2008, based on a call for improved mathematics instruction by the Idaho State Department of Education, the Idaho legislature mandated that all teachers who might potentially teach mathematics, as well as administrators (about 11,000 people), take the MTI course for recertification by 2014. The pool of MTI participants includes all teachers in the state with a K–8, special education, or secondary mathematics certification, and all those with administrator certifications. Although the course is offered at three levels with largely overlapping content (early elementary, intermediate, and secondary), a single class may include individuals with a very broad spectrum of mathematics backgrounds. It is not uncommon, for example, to have a middle school English teacher, a high school calculus teacher, and a district administrator all in the same secondary course. The 45-hour course focuses on early number, rational number, operations, and algebra. More detail on the development and implementation of the course can be found in Brendefur, Thiede, Strother, Bunning, and Peck, 2013; Brendefur, Carney, Hughes, and Strother, 2015; and Carney, Brendefur, Thiede, Hughes, and Sutton, 2014.

Changes in educators' mathematical content knowledge were measured precourse and postcourse, using items around number and algebra from the Learning Mathematics for Teaching project at the University of Michigan (Learning Mathematics for Teaching, 2008). In addition, retrospective changes in teachers' beliefs about mathematics and self-efficacy in teaching mathematics were measured postcourse with a survey. (Survey items were designed by RMC Research and Math in the Middle project staff at the University of Nebraska–Lincoln in 2005.) For example, participants were asked whether they agree, disagree, or neither agree nor disagree with the statements "All students can learn challenging content in mathematics" and "Mathematics should be learned as sets of algorithms or rules that cover all possibilities." Items were administered retrospectively postcourse in order to avoid pre- to post-intervention response-shift bias, in which participants' understanding of the survey items changes in response to the intervention (Bray, Maxwell, & Howard, 1984), in this case the MTI course. For example, prior to the course, many participants were unfamiliar with the word *algorithm* and may have based their definition of "challenging content" solely on procedural complexity rather than conceptual depth. Using a retrospective presurvey/postsurvey design has been documented as addressing this issue (Aiken & West, 1990; Lam & Bengo, 2003).

Data from the first 3 years of the MTI project show statistically significant gains in participants' content knowledge as well as significant shifts in their beliefs about the nature of mathematics and how it should be taught (Carney, Brendefur, Thiede, Hughes, & Sutton, 2014). These initial results led us to think more deeply about what aspects of the course and its facilitation might be responsible for shifts in participants' content knowledge and beliefs. In this article we analyze participant work, course transcripts, and evaluation feedback around the facilitation of a single task that exemplifies the overall course structure.

MTI Course Theoretical Background

The MTI course uses social-constructivist learning theory and was built around the concepts of teaching for understanding as described by Hiebert and others (Hiebert & Carpenter, 1992; Hiebert, 1997), in which learners construct their mathematical knowledge through their own mathematical activity and interactions with their peers and their teacher. More specifically, the course seeks to build teachers' mathematical knowledge for teaching through a series of tasks that require them to take part in the process of progressive formalization based on the Dutch concepts of "guided reinvention" and "realistic mathematics education" (Freudenthal, 1973; Freudenthal, 1991; Gravemeijer & Doorman, 1999; Treffers & Vonk, 1987). In guided reinvention, students are given an experientially "real" task that allows them to use the mathematics they do know as an entry into a new mathematical concept (horizontal mathematization). From these initial strategies, students then move through tasks, classroom discussion, and instruction to move toward more efficient and abstract mathematics (vertical mathematization).

Guided reinvention can still take place in professional development interventions because many teachers have memorized algorithms but have little understanding of why they work. When pressed to solve tasks without those algorithms, teachers with limited conceptual understanding will go through the process of reinventing mathematics, albeit in an accelerated form, connecting their own initial informal methods to progressively more formal methods introduced by their peers and instructor. Teachers with a more conceptual mathematical background are pressed to consider what initial informal methods might look like and how they connect to more formal methods.

The actual facilitation of the MTI course largely mirrors the orchestrating classroom discussions framework, which consists of (1) anticipating how students will approach a task; (2) monitoring students' progress, in part through

questioning; (3) selecting students' solutions for discussion in order to highlight specific mathematical ideas; (4) sequencing the student work in order to press connections; and (5) connecting across strategies to highlight key ideas and connect to formal mathematics (Smith, Hughes, Engle, & Stein, 2009; Stein, Engle, Smith, & Hughes, 2008). Participants in the MTI course are also asked to solve or examine a given task by creating and connecting between enactive representations (physically constructing or acting out the situation), iconic representations (visually modeling the situation with a drawing or visual tool such as a number line), and symbolic representations (numbers, variables, tables, and equations) that characterize various levels of formality (Bruner, 1964).

The Developing Mathematical Thinking Framework

The MTI course highlights five central practices for developing mathematical thinking that combine socio-constructivist theory and progressive formalization: (1) take students' ideas seriously, (2) press students conceptually, (3) encourage multiple strategies and models, (4) focus on the structure of mathematics, and (5) address misconceptions (Brendefur, 2008). This framework forms the basis of both our work with teachers in their classrooms as it pertains to students and for professional development with teachers. Each practice is discussed briefly below; detailed descriptions can be found in related publications (Brendefur et al., 2015; Carney et al., 2014).

Take students' ideas seriously. Of centrality is the idea that instruction should be based on student thinking. By monitoring students' approaches to a mathematical problem, the teacher determines what students know, what misconceptions they may have, and the breadth of understanding or experience across the class. One must also anticipate students' strategies so they can be quickly identified during the monitoring process. In addition, by selecting student (or participant) representations for class discussion, students can take ownership of the mathematical knowledge that emerges.

Press students conceptually. Once students (or participants) have applied their own thinking to a problem, the teacher must press them to generalize and abstract. This can largely be accomplished by pressing students to make connections between different student representations or solution strategies and by further connecting those representations to more formal concepts or representations. In this way, students gain familiarity with formal mathematics but see its direct connection to their own initial mathematical knowledge. This reflects a cyclic process of guided reinvention whereby students are pressed from their initial thinking to invent more formal efficient methods, which in turn become future initial strategies.

Encourage multiple strategies and representations. In the process of solving a rich contextual problem, a class may generate several solution strategies and ways of representing them. The course stresses that formal algorithms are simply one strategy for solving problems and that working with multiple approaches and enactive, iconic, and symbolic models strengthens students' understanding of the mathematics.

Focus on the structure of mathematics. There are fundamental mathematical ideas that run throughout mathematics. By focusing on these structural ideas and connecting them to student solutions and representations, teachers ensure students make connections across their mathematical experiences, which helps them become problem solvers. For example, decomposing numbers (e.g., 5 equals 3 + 2) is necessary for developing derived facts strategies and for seeing equivalent forms of algebraic expressions (e.g., $5x$ equals $3x + 2x$). Tricks, such as the "add a zero" rule for multiplying by 10, circumvent structural mathematical understanding and can thus be forgotten or misapplied. The MTI course intends to increase teachers' awareness of these structural components of mathematics.

Address misconceptions. We frequently observe deep misconceptions that stem from a lack of structural mathematical understanding. For example, a student who performs $1/2 + 1/3 = 2/5$ is missing a central understanding of what fraction notation indicates and possibly the deeper idea of "unit." This misconception would traditionally be addressed by simply restating the algorithm for adding fractions. The MTI course addresses how such mistakes indicate deeper mathematical misunderstanding and need to be addressed at a foundational level.

These five practices for developing mathematical thinking are central to the facilitation of the entire course, particularly in the division of fractions task described in this article, called the Maren's Garden problem. The Maren's Garden problem was also chosen because comments from participants and related feedback on the course evaluation (discussed under Implications: A Framework for Tasks that Exemplify Guided Reinvention) led us to consider that the attributes of this particular task may be useful for those creating mathematics professional development courses.

Task Facilitation

Prior Work

Prior to this task, participants spend 20–30 hours investigating number, operations, and algebra. Precursors relevant to this task include the following.

1. Understanding the power of context when making sense of operations. Participants investigate how computation algorithms develop from the need to model real situations and are based on foundational properties of numbers.
2. Developing multiple models for division from informal solution strategies.
3. Modeling fractions and fraction operations with number lines, bar models, and Cuisenaire® rods.
4. Investigating the following three concepts foundational for number understanding in general and fraction work in particular.
   - *Iterating and partitioning:* Iterating refers to the process of repeating a single unit, such as counting by thirds up to four, whereas partitioning refers to the opposite process of breaking into equal-sized pieces.
   - *Units and unitizing* (Lamon, 1999): Fluently and flexibly changing the unit being considered. For example, conceptualizing and operating with two 1/3 pieces as a new unit of 2/3.
   - *Equivalence and relationships:* Comparing and equating rational numbers, such as determining whether two fractions are equal.

These concepts overlap substantially in practice. For example, one might change the units of a given quantity by partitioning the unit (e.g., moving from fourths to eighths) in order to establish equivalence to another quantity. The Maren's Garden task is generally part of the culminating set of activities for this topic. This task would address fifth-grade standards in the CCSS around division of a whole number by a fraction and using visual representations and equations to solve real-world problems.

The Maren's Garden Task

The task is presented to participants as follows:

> Maren ordered 4 bags of soil for her raised flower gardens. Each garden needs three-fourths of a bag of soil. How many gardens can she fill completely with soil? How much soil does she have left?[1]

Participants are asked to solve the problem with two different methods of their choosing, but at least one method must incorporate a visual model. The answer to the task is 5 gardens with 1/4 of a bag of soil left over, but the traditional multiply by the reciprocal algorithm yields 5 1/3. This discrepancy is intentional and leads to cognitive dissonance and engagement as seen in the participant solutions below.

Video of the Maren's Garden problem facilitation was taken for two MTI courses. Segments of this video thought to exemplify participants' responses to the problem were transcribed, and pictures of participants' work are used where possible. Participant names have been changed.

Initial Participants' Solutions

Participants tend to approach the problem in one of two ways. Many begin by drawing a visual representation (informal iconic) of the four bags, divide them into fourths, and count how many groups of the three 1/4 pieces there are, as shown in participant Jane's work (Figure 1).

Participants using this method generally arrive at the answer of 5 with 1/4 left over. In Transcript 1 Jane describes her representation to the whole class.

Transcript 1

Jane: I just had four bars there; each bar represented a bag. Then I divided each bag into fourths. I thought, OK, a garden would take—first garden [gesturing to leftmost bar] would take up the first three [pointing to vertically striped boxes], then I did different pattern for the second garden, then I just kept on taking threes, um, until I got to where I couldn't take any more threes. And then went back and found I had one, two, three, four, five gardens with one quarter of a bag left.

1 This task was first used in a professional development workshop funded by a Mathematics and Science Partnership grant for Brendefur in 2004.

Figure 1. **Jane's picture—showing gardens in bags via different shading.**

Of particular note is that she begins to call the three 1/4 bags that make up a garden just "three," dropping the unit of "1/4 bag" altogether until the very end of her explanation. In essence, she has turned the problem into

(1) $$4 \div \frac{3}{4} = \frac{16}{4} \div \frac{3}{4} = 16 \textit{ quarters} \div 3 \textit{ quarters} = 5\ R\ 1$$

where the remainder is in the unit of quarter bags. Thus, in participants' answer of 5 and 1/4, the 5 refers to gardens, whereas the 1/4 refers to bags. Most participants, however, simply see their solution in the picture itself rather than as this common denominator algorithm.

Participants using more traditional mathematics tend to jump to the algorithm for solving 4 divided by 3/4, multiplying 4 by 4/3 and ending with an answer of 5 1/3.

(2) $$4 \div \frac{3}{4} = 4 \times \frac{4}{3} = \frac{16}{3} = 5\frac{1}{3}$$

Because we require a visual model, many participants discover this discrepancy within their own work, while others identify it within their group.

Lines of Questioning

Depending on their thinking and models, groups are then presented with a series of questions and prompts.

1. Why is there a discrepancy between the answer from the picture and the answer from the algorithm?

Subquestion: Represent the problem with one of the other models for division or fractions used in the course (e.g., Cuisenaire rods, double number line, ratio table, partial quotient).

Subquestion: What part of a garden could you make with the leftover soil? How much of another bag would you use to fill another garden?

2. How does the traditional algorithm make sense in context? What does 4/3 mean in the context of the problem, and why does it make sense to multiply it by 4?

Subquestion: If you have created a ratio table that starts with 1 garden = 3/4 bags, create a ratio table that starts with 1 bag =__ gardens.

Subquestion: If you have created a picture or double number line, where can you see the 16/3?

These questions have three goals with respect to teachers' content knowledge. First, we want participants to make the connection that 1/4 of a bag of dirt is equivalent to 1/3 of a garden and thus realize that the unit one chooses to operate with is flexible. Second, they should be able to articulate that the "multiply by the reciprocal" algorithm is operating with the inverse of the original unit, changing from dividing by bags per garden to multiplying by gardens per bag (equation 2). And lastly, participants should be able to develop a common denominator algorithm for fraction division (equation 1) by recognizing that in many of their strategies they are converting the dividend into the same units as the divisor, namely quarters. Pedagogically, we hope that this process of questioning groups, whole-class discussion, and making connections between different representations takes teachers through the process of guided reinvention as though they were students, demonstrating the Developing Mathematical Thinking framework.

The transcript below exemplifies the line of questioning the instructor might take once a participant sees the discrepancy between the leftover 1/4 and 1/3 but is unsure about why it exists. We use symbolic versions of numbers when people are referring to values within an equation and express numbers as words when people refer to those values more within context. The instructor here begins by asking the participant to reiterate what each quantity means within the context of the problem.

Transcript 2

Instructor: [Referring to the numbers written on the participant's page] What is the 5; what is the 1/4; what is the 3/4; what is the 1/3?

Sue: 5 is the number of gardens we are filling . . . 1/4 is what was left out of the bag, out of four bags of soil . . . 3/4 is how much was required for each garden . . . So 1/3 is . . . hmmm . . .

Molly: A third of this [pointing to the 3/4]? Of the 3/4?

Sue: 1/3 is what is actually left in the bag, right?

Instructor: OK, so go back to your drawing here [similar to Figure 1]. Where's the 3/4?

Sue: Right here [pointing to one of the gardens in the picture].

Instructor: OK, so that's this [pointing to the 3/4]. What is this [pointing to the 1/4]?

Sue: 1/4 is this [again pointing to leftover quarter in the picture], what's left over after . . .

Molly: There's just a quarter left out of the bag.

Instructor: OK, let me do it this way then. Take that last quarter that's left. You have a quarter left. Throw it into the last garden. What's happened to that garden?

Molly: Too much soil.

Instructor: No, you just have . . . how much soil do you have left? A quarter . . . and you throw that into the last garden. How much of the garden is full?

Sue: Only a third of the—if we were to have another garden . . .

Molly: 1/3 of our sixth garden!

The instructor uses physical imagery (throwing soil into an additional garden) to press the participant to make a connection between the two different units of bags and gardens. A common tactic among instructors, as illustrated above, is to bring the participant's attention back to the context of the problem and the numbers related to that context. This tactic is meant to both help participants who have difficulty making sense of the numbers and operations they have written down and to press more advanced participants to look beyond known algorithms.

Additional Representations

Facilitation of this task relies on participants generating and connecting between multiple representations of the problem in order to meet the three goals described above. We therefore examine participant work for each of the two most common representations aside from the traditional algorithm and informal picture.

Double number line. Instructors often ask questions encouraging participants to translate between units. The double number line shows equivalence between units in proportional relationships and thus can be used to model multiplication and division. It is also called for in the CCSS in sixth grade as part of proportional reasoning. Jim created the number line in Figure 2.

Transcript 3

Jim: We had the four bags on a number line, uh, since it was three fourths of a bag per garden, we divided the bags into fourths, then just counted basically 3/4, 3/4, 3/4, 3/4, and 3/4 and got five gardens. And, uh, it was just the one fourth of a bag left over.

Instructor: And how do you see that that one fourth is also one third of a garden?

Jim: Well, uh, this unit [motions to the leftover quarter] is the same as this unit right here [motions to first quarter bag on the line], which was one out of the three that was included in the bag.

In this model, equal length indicates equal value. Jim's double number line indicates how many bags of soil are

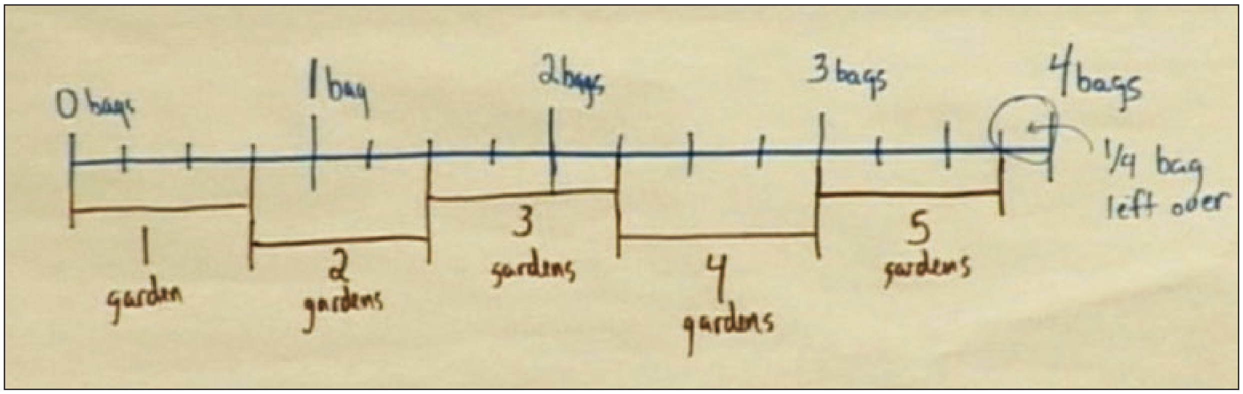


Figure 2. **Jim's double number line representation**

in a given number of gardens and vice versa. Of particular interest is the interpretation of the final quarter bag. One can compare the length of the 1/4 bag to the 3/4 bag required to create one garden. That 1/4 bag length is thus equivalent to 1/3 of a garden.

Ratio tables. Tables of equivalent ratios (or ratio tables) are used to model multiplication or division and to examine proportional relationships. They are explicitly called for in the sixth grade CCSS. In this problem, two different ratio tables can be used to model the problem (Figure 3). Ratio table (a) models the problem as it is given—the number of bags per garden. But ratio table (b) examines how many gardens one could make from each individual bag.

Sasha was asked to relate each of her two ratio tables back to the unit in which she is operating. In doing so she (at this point, unknowingly) is describing the two different strategies for dividing fractions.

Transcript 4

Sasha: So I did my ratio table both ways. So in the first one [pointing to table (a)] G is gardens, and B is bags. And then the arrows—this is still the bags column, I just wrote it so we can see it as 1 1/2 garden . . . so I mean bags. 1 garden is 3/4 of a bag, 2 gardens is 1 1/2 bags, 3 gardens is 2 1/4 bags, so you get down to 5 gardens is 3 3/4 bags. . . . So you see that we have 5 gardens and we had to use 3 and 3/4 bags, so we have 5 gardens and a quarter bag left over. And then going this way [pointing to table (b)], if you say bags, you have 1 bag, that's 4/3 of a garden, which is really 1 1/3 gardens, 2 bags is 2 2/3; 3, 4 gardens, so I have 4 bags, is 5 1/3 gardens, so I have five gardens and then a third of a garden left over.

Connecting representations. Facilitators model the practice of connecting between models to press participants conceptually toward more abstract or efficient methods while connecting back to less formal representations. For example, in this transcript, the participant initially explains the ratio table (Figure 3), but then the facilitator asks him to relate the ratio table to the number line (Figure 2).

Transcript 5

Sam: So that's 12/4, which is 3, right. So—and 4 = 3 [pointing to the 4 bags soil to 12/4 garden in the ratio table]. So then if we go to 5, if we want 5 bags, then we have 3 and . . . it's like doing that counting... so it's 3 3/4. So that means you have a quarter left. It's like counting up by 3/4 and counting down by 3/4. So this would be—2 would be 1 1/2, 3 would be adding another 3/4 again, and that would be 2 1/4. Then it's 3, and then it's 3 3/4 on the ratio table.

Instructor: How does that relate to your number line there?

Sam: I counted up . . . I took a bag, 3/4, and I made 3/4 be 1. Another 3/4 is a unit 2. Another 3/4 is a unit 3. Another 3/4 is a unit 4. These were left over, and so I just moved one over there and got to 5. And this was 1/4 left. On the number line, I made the units be—each unit be 3/4.

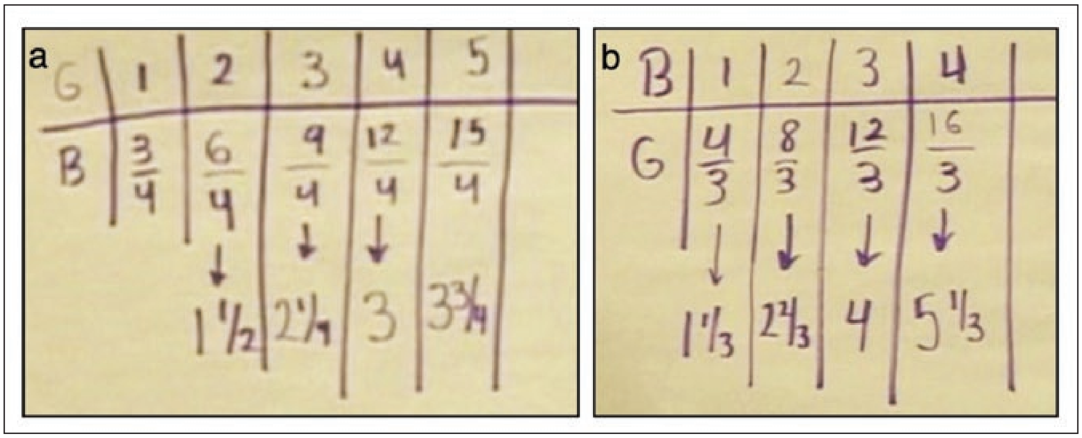


Figure 3. **Sasha's representation—ratio tables a) counting in units of gardens; b) counting in units of bags.**

Developing Algorithms

Figure 4 shows two *possible* progressions of formalizing through multiple representations, indicating how different strategies might build to two formal algorithms.

Multiplying by the reciprocal. The left column of Figure 4 connects the traditional algorithm for fraction division to the other representations. The 3/4 in 4 / 3/4 represents 3/4 bag per garden. The reciprocal can be thought of as either 4/3 gardens per bag or 4 gardens for every 3 bags. Both interpretations are evident in the ratio table and visual representations, where each bag contains 1 1/3 gardens' worth of soil. Once participants have shared their models, the instructor generally asks a participant to explain the connection between the visual models and the traditional algorithm, as exemplified by the following transcript where, during whole-class discussion, Chris connects the traditional algorithm to the ratio table in Figure 3b.

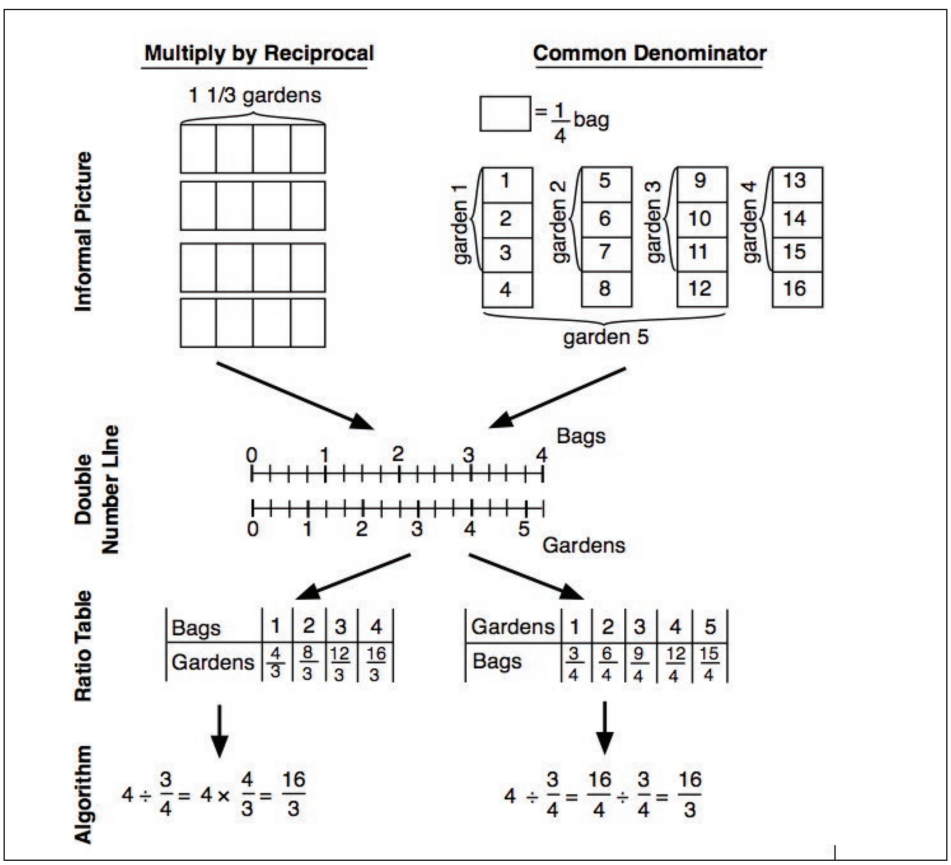


Figure 4. **Possible ways of progressively formalizing toward fraction division algorithms.**

Transcript 6

Instructor: Where is this 4 times 4/3 up here [gesturing to representations on the board]?

Chris: It's right here [pointing to first column of Figure 3b]. 4/3 times four to get . . . 4/3 of a garden out of one bag. 4/3 times 4 bags gives you 16/3 [pointing to fourth column], which is 5 1/3. So this 4/3 [pointing to 1:4/3 in ratio table] is this 4/3 [pointing to 4/3 in traditional algorithm].

Common denominator algorithm. Both the double number line and the informal picture bring out a second division algorithm, as seen in the right column of Figure 4. In this problem, we can view the four bags of soil as containing a total of 16 quarter-bags of soil. Viewing quarter-bags as our unit, we then measure out groups of three units to determine how many gardens can be made. There are five groups of three or five full gardens with one unit left over. We are changing all parts of the problem into the same fractional unit, allowing us to essentially ignore the unit altogether. In this case, 4 divided by 3/4 becomes 16/4 divided by 3/4, which we can treat as simply 16 divided by 3, as in Equation 1.

Technically, one need not find common denominators and could instead simply divide straight across in the same manner in which we perform fraction multiplication. But in even fairly simple problems, this can result in ugly fractions that have to be extensively manipulated

$$\left(\text{e.g.,} \frac{3}{4} \div \frac{2}{3} \text{ ends up being } \frac{3/2}{4/3}\right).$$

Teachers seldom develop the common denominator algorithm on their own during the task facilitation. Rather, after the problem has been largely debriefed, the instructor will put up the method on the board, asking teachers whether it is mathematically valid and what it represents within the problem context and the multiple representations. Some teachers do, however, develop the 16 units on their own within the drawing (Figure 1), as this transcript indicates.

Transcript 7

Tammy: This is my area model here, so I decided—so these are all bags. Bag, bag, bag [pointing to her drawing, similar to Figure 1]. I divided each bag into four. We have 16 total parts of soil. Each garden takes three parts, so I just decided to divide these up into three. Each shaded area is a garden . . . so there's a garden, there's a garden, there's a garden, there's a garden, there's a garden. So there's my three parts and then this little guy's left over [pointing to the unshaded quarter garden]. So that's 1/16 of soil left over. Or if I had two more pieces of my area I could fill a garden. So it's 1/3 of a garden completed.

Instructor: OK, and when you call it 1/16 . . .

Tammy: Parts of soil.

Instructor: Parts of? What is your whole then?

Tammy: 16 . . . my whole is four bags of soil, which is 16 parts of soil if I divided it into four. So I have 16 total parts.

Tammy is seeing the problem as four bags divided into sixteen equal pieces, each of which represents one fourth of a bag. This recognition is essential for developing the common denominator algorithm.

Course Evaluation Data

In order to capture participants' thoughts about this task, we analyzed their responses to three evaluation questions for courses taking place during the spring and summer semesters of 2013, pulling those responses that mention this task specifically (either by name or as a division of fractions task). This time frame was chosen for convenience because at this time course evaluations moved online, making it easier to filter participant responses. The specific qualitative evaluation items reported in this study are:

1. Describe the activities you found most useful and why (please be specific).
2. What new mathematics did you learn during the course? Comment on any mathematical "aha" moments you experienced.
3. Describe any new teaching practices you learned (specific to mathematics) and how you intend to use them in your classroom.

Although the course evaluation responses represent a temporal snapshot of the 5-year administration of the MTI course and were made anonymously, we consider them to be representative of the course participant population. This sample includes participants from all three levels of the course (797 early elementary, 417 middle grades, 120 secondary).

Of the 1,145 participants who responded to at least one of the three evaluation questions of interest (about 60% of all participants), 27% mentioned the fractions unit in one of their responses, and 59 people specifically mentioned the Maren's Garden problem. The fractions unit represents approximately one eighth of the course. Given the open-ended nature of the questions and the fact that over a quarter of responding participants mentioned the fractions unit generally and 5% specifically mentioned this problem, the tasks contained within seem to have been particularly impactful. Quotes from the evaluation questions in which participants specifically cite the Maren's Garden problem are used to support and exemplify the attributes of the task that have implications for mathematics professional development more generally.

Implications: A Framework for Tasks that Exemplify Guided Reinvention

Based on evaluation responses and the authors' collective experience in facilitating the task over 90 times, we present a framework of factors that make this activity useful in exemplifying the process of guided reinvention, even for participants who know a formal algorithm. Each aspect of the framework is accompanied by quotes from participant evaluations in order to illustrate the factor from the participant's perspective (the number following the quote refers to which of the three above questions the participant was answering).

Authentic Context, Visual Situation, and Multiple Solutions

This problem can be solved with a variety of models because the context lends itself to visual representation. In addition, all of the different representations connect to one another (Figure 4). This allows teachers with more limited math backgrounds to feel validated in their initial reasoning while pressing more mathematically confident teachers to find their algorithm within the less formal solutions. It is through connecting representations that participants take part in progressive formalization firsthand. The idea that problems can be approached informally seems to resonate strongly with many teachers.

Quote 1: Being able to multiply and divide fractions without using an algorithm (because of course I don't remember them) and getting the right answer by using an iconic model and actually understanding what I did and why it worked!! (Question 2)

Quote 2: Why the traditional algorithm for division of fractions requires that you use the reciprocal. It opened me up to the realization that not everything needs a computation, sometimes logic thinking is best. (Question 2)

In addition, this task lends itself to developing multiple algorithms because of its structure. This is a measurement division task for which two different broad solution methods are possible (Figure 4). We found partitive fraction division contexts to be less fruitful for initially developing multiple strategies because they tend to lead to the reciprocal algorithm. For example, questions of the form

> If 4 bags of soil fill 3/4 of a garden, how many bags are required to fill one whole garden?

tend to lead participants toward iterating 4/3 four times. Thus, selecting problems for which multiple informal strategies could be used allows for greater flexibility in how the mathematics is reinvented.

Cognitive Dissonance

In professional development teachers' time with a given topic is likely limited to a few hours, and engaging a wide range of teachers in the same problem can be difficult because some have formalized content knowledge and some do not. Given these constraints, we seek to ensure that everyone has access to the problem and can be cognitively engaged. Intentionally creating cognitive dissonance is one way to ensure that teachers are pressed beyond just "doing" the problem.

In this case, the multiple solution paths lead to two seemingly different answers: 5 1/3 (both in the unit of gardens) or 5 and 1/4 (where the 5 is the unit of gardens but the 1/4 is in the unit of bags). This creates temporary confusion that generates discussion and questions amongst participants. This dissonance is particularly valuable in generating discussion between participants who drew informal pictures and those who used an algorithm because their answers appear not to match—but both camps are confident in their approach. It is this apparent dissonance in a professional development setting that can help illustrate the value of discourse as participants slowly realize the commonalities and connections between their solutions.

Quote 3: When we were solving Maren's Garden I wanted to go directly to the standard algorithm but I drew my picture and then solved the problem with the standard algorithm. The answers didn't match so I was trying to figure out what I did wrong that made the answers different. It was when someone put a different visual on the board that I realized why drawing a picture gives you a different answer than the algorithm. Then for fun we got

the fraction rods out and solved the problem with those and saw an even better visual. (Question 2)

Quote 4: It was very eye opening to see that the algorithm's fraction did not match the picture model and a number line was used to understand the meaning of the fractions. (Question 2)

Deep Engagement

Because of the multitude of representations and the cognitive dissonance, this task typically takes about 1.5 hours to facilitate. This extended focus on a single problem may encourage teachers to rethink the nature of mathematics and its instruction, requiring them to see the development of number operations as a problem-solving opportunity in and of itself and, more broadly, what it looks like to do mathematics rather than computation.

Quote 5: I kept dreaming last night about 3/4 of a bag of soil per garden and that 1/4 is left over to cover 1/3 of another garden and so on . . . and this dream was neverending! My mind is exhausted! But I'm really learning a lot and have been trying to teach math from a different point of view now! It's been great! Thanks! (From uninitiated email correspondence with a participant, reported here with permission.)

Changing Mathematical Knowledge

This activity in particular focuses on both building content knowledge and changing knowledge about mathematics and how it is taught. The activity is couched within a series of smaller, content-specific tasks that build teachers' conceptual understanding of fractions, their representations, and the more general concepts of units and unitizing (see **Appendix A** for the unit overview). We hypothesize that our professional development would be far less effective if it consisted only of small tasks that focused on specific concepts or, conversely, only included extended problem-solving tasks such as the Maren's Garden problem. It is this combination and the sequencing of tasks that allows participants to model the problem themselves (nontraditionally) and connect between models, thus building both their mathematical content knowledge and pedagogical content knowledge. We see both of these ideas in the quotes below.

Quote 6: I had some "aha" moments with multiplying and dividing fractions. I was able to see how they are connected to real world situations. I learned the algorithms and was able to compute accurately but could not understand what was happening. Thanks to MTI I have a better understanding. (Question 2)

Quote 7: In my work with classroom teachers, I intend to really push them on de-emphasizing a specific method and taking the time to develop the concept through student-developed (teacher influenced) models. That was probably my greatest take away from this training. . . . It really does take a lot of time to develop a deep understanding and it is important that teachers take the time. (Question 3)

Implications for Teacher Educators

In planning for work with preservice or in-service teachers, we suggest that mathematics teacher educators might use these four attributes in evaluating potential tasks. Problems from curricula that teachers are familiar with can be used to reinforce content knowledge but may lack multiple points of entry or the cognitive dissonance that deeply engages adults and has the potential to change their perceptions about mathematics. But one might alter such problems to ensure that multiple representations are possible by changing the number set (e.g., from large values to whole numbers or simple fractions) or changing the context to elicit specific visual representations (e.g., from money to distance or area). Likewise, one can foster cognitive dissonance by altering the task such that there is a single solution that has multiple, seemingly different, forms. Removing directions such as "Create an equation that . . ." or "Use a table to . . ." can give the flexibility to approach the problem in different ways, leading to different methods and solution forms. Conversely, mathematics professional development sometimes introduces advanced topics such as combinatorics or group theory in tasks that are useful in helping teachers develop problem-solving strategies. But we have found that using familiar mathematical content can help teachers make strong connections to their own practice, resulting in changes in their mathematical knowledge for teaching.

Conclusion

The goal of this article is to provide a detailed analysis of a single professional development task that engages teachers in mathematical discourse that can challenge and change their specific content knowledge. But based on participants' comments and reported changes in beliefs, such tasks may also change teachers' perceptions about the nature and teaching of mathematics itself by allowing them to take part in the activity of guided reinvention by using context and models to develop and make sense of computation algorithms. At a time when teachers are seeking opportunities to increase content knowledge, we argue that professional development focused on mathematics content must also change

foundational beliefs about mathematics, pressing teachers to see mathematics as a sense-making activity. It is notable that this task is relatively simple; it does not require a great deal of setup, uncommon materials or manipulatives, or any technology. Rather, it is through the facilitation and questioning that deep mathematical connections are made. This problem and problems that have similar traits (e.g., multiple representations and cognitive dissonance) might thus be used in a wide variety of professional development settings to increase content knowledge while modeling how the teaching of computation can be utilized as a problem-solving activity.

References

Aiken, L. S., & West, S. G. (1990). Invalidity of true experiments: Self-report pretest biases. *Evaluation Review, 14*(4), 374–390.

Ball, D. L., & McDiarmid, G. W. (1989). *The subject matter preparation of teachers*. East Lansing, MI: National Center for Research on Teacher Education.

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

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3–15.

Bray, J. H., Maxwell, S. E., & Howard, G. S. (1984). Methods of analysis with response-shift bias. *Educational and Psychological Measurement, 44*(4), 781–804.

Brendefur, J. L., Carney, M. B., Hughes, G., & Strother, S. (in press). Framing professional development that promotes mathematical thinking. In E. Ostler (Ed.), *Emerging trends and perspectives in STEM learning*.

Brendefur, J. L. (2008). Connecting elementary teachers' mathematical knowledge to their instructional practices. *The Researcher, 21*(2), 1–18.

Brendefur, J. L., Thiede, K., Strother, S., Bunning, K., & Peck, D. (2013). Developing mathematical thinking: Changing teachers' knowledge and instruction. *Journal of Curriculum and Teaching, 2*(2), 62–75.

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist, 19*(1), 1–15.

Carney, M. B., Brendefur, J. L., Thiede, K., Hughes, G., & Sutton, J. (2014). Statewide mathematics professional development: Teacher knowledge, self-efficacy, and beliefs. *Educational Policy*, prepublished September 21, 2014. doi:10.1177/0895904814550075

Ernest, P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A model. *Journal of Education for Teaching, 15*(1), 13–33.

Fosnot, C. T., & Dolk, M. (2002). *Young mathematicians at work: Constructing fractions, decimals, and percents*. Portsmouth, NH: Heinemann.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, the Netherlands: Reidel.

Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Dordrecht, the Netherlands: Kluwer.

Gardner, D. P. (1983). *A nation at risk*. Washington, DC: U.S. Government Printing Office.

Gravemeijer, K., & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics, 39*(1), 111–129.

Gravemeijer, K., & van Galen, F. (2003). Facts and algorithms as products of students' own mathematical activity. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to "Principles and standards for school mathematics"* (pp. 114–122). Reston, VA: National Council of Teachers of Mathematics.

Hiebert, J. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York, NY: Macmillan.

Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education, 35*(5), 330–351.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.

Lam, T. C. M., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation, 24*(1), 65–80.

Lamon, S. J. (1999). *Teaching fractions and ratios for understanding: Essential content knowledge and instructional strategies for teachers*. Mahwah, NJ: Lawrence Erlbaum.

Learning Mathematics for Teaching. (2008). *Mathematical knowledge for teaching items: Number, operations, and algebraic knowledge of students and content*. Retrieved from http://sitemaker.umich.edu/lmt/home

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well-taught" mathematics courses. *Educational Psychologist, 23*(2), 145–166.

Smith, M. S., Hughes, E. K., Engle, R. A., & Stein, M. K. (2009). Orchestrating discussions. *Mathematics Teaching in the Middle School, 14*(9), 548–556.

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning, 10*(4), 313–340.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap*. New York, NY: Free Press.

Treffers, A., & Vonk, H. (1987). *Three dimensions: A model of goal and theory description in mathematics instruction—The Wiskobas Project*. Dordrecht, the Netherlands: Reidel.

Authors

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

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

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

Appendix A: Fraction Unit Sequence for MTI Course

The sequence of activities below represents 8–12 hours of coursework in the 45-hour class.

1. Defining fraction
   a. Enumerating a denomination
   b. Part–whole interpretation
2. Connecting to whole number
   a. Composing and decomposing fractions to add and subtract
   b. Skip counting by fractional units
3. Iterating and partitioning on the number line
4. Unitizing within a bar model and connection to fraction multiplication
5. Unitizing within an area model
6. Defining unit and “whole” with Cuisenaire rods (e.g., if a given color = 1, what do other color rods represent)
7. Maren’s Garden task (division of fractions)
8. Multiplication of fractions in an area model
9. Analyzing student strategies for subtraction of mixed numbers

(**Return to page 126**)

Openness and Measurement: Two Principles for Improving Educational Practice and Shared Instructional Products

Anne K. Morris
University of Delaware

James Hiebert
University of Delaware

Two studies were conducted to identify the conditions under which instructors teaching the same mathematics teacher preparation course would continuously improve their shared instructional products (lesson plans for class sessions) using small amounts of data on preservice teacher performance. Findings indicated that when lesson-level student performance data were simply collected, by course section, the instructors could make important changes to the lessons but did not often do so. However, when the instructors were encouraged to compare data across semesters, they generated hypotheses that guided instructional improvements, which then were tested through multiple cycles. The cycles of hypothesis testing helped instructors clarify the goals for improvement, use the performance data to test whether changes were actually improvements, and reduce their tolerance for marginal student performance.

Key words: Clicker technology; Continuous improvement; Improvement science

For more than a decade, mathematics educators at our university have been working on a system of data-based continuous improvement for preparing K–8 mathematics teachers (Hiebert & Morris, 2009). For us, this means improving incrementally, each semester, the effectiveness of the four courses we offer for preparing preservice teachers to teach mathematics well to K–8 students. In particular, this means continuously improving the shared lesson plans—*the shared instructional products*—used by all instructors of the courses (Morris & Hiebert, 2011). Although there is evidence that these efforts have produced increased preservice teacher learning (Austin, 2012; Berk & Hiebert, 2009), we also have become painfully aware of the challenges of developing a sustaining system of improvement based on student performance data. To develop such a system we must learn how to help instructors use small amounts of systematically collected data to make and test important changes to the lessons.

Why do we think using small amounts of data to revise the lessons is the key to a sustaining system of improvements to the course lessons? We can identify two reasons. First, although controlled experiments have been carried out to test the effects of some changes to the lessons (e.g., Berk, Taber, Gorowara, & Poetzl, 2009), these experiments are expensive in time and effort. We cannot depend on these expensive studies to sustain a continuing improvement process. Collecting and analyzing small amounts of data is more likely to fit into the weekly work lives of instructors making the activity sustainable while also providing just enough data to test whether changes to lessons are promising improvements. Second, the model we have developed and employed to initiate the system of course improvement has been successful; but over time, we have noticed one missing ingredient—the regular collection of small amounts of data to guide proposed changes to lessons.

The model that leads to continuously improving instructional products (Morris & Hiebert, 2011) has three key features: (1) working toward the same learning goals, (2) eliciting ideas for improvement from multiple sources, and (3) using small amounts of data to repeatedly guide and test changes. In our setting, the learning goals are shared across instructors and across time because they are specified in the lesson plans used to teach each class period during the semester. Ideas are elicited from multiple sources by searching the literature for approaches to teaching preservice teachers and by soliciting ideas from new instructors (faculty and doctoral students) from year to year. But we have noticed that changes often are made to the lessons based on a rational analysis of the lesson or by anecdotes shared by instructors drawn from the responses of a few vocal students during the class session. The changes prompted by these analyses often seem to be minor changes that do not substantially improve the effectiveness of the course. How can we move instructors

Preparation of this article was supported by the National Science Foundation (Grant #0083429 to the Mid-Atlantic Center for Teaching and Learning Mathematics). The opinions expressed in the article are those of the authors and not necessarily those of the Foundation.

to use data on student performance instead of their own logic and anecdotes to improve the lessons?

Two Promising Principles: Openness and Measurement

Under the term *Improvement Science*, the quality improvement leaders and practitioners in clinical medicine and business have been studying for some time how to use small amounts of data to improve services and products (Berwick, 2003, 2005, 2008; Gawande, 2007; Kenney, 2008; Langley et al., 2009; Rother, 2009). We borrowed two principles from these fields that have led to incremental but striking improvements in the quality of outcomes: openness and measurement.

Openness means allowing one's professional practice to be open for inspection. Openness shifts professional practice from a private practice and an indicator of personal competence to a public opportunity for learning. Not just any kind of openness is productive. Learning seems to occur as data about practices and outcomes are compared and studied to identify those that most effectively achieve the desired goals.

The positive effects of openness are dependent on sharing reliably measured information on practices along with the outcomes of those practices. Comparing and studying practices requires measuring key features and outcomes of practices, not just sharing stories or anecdotes. Deciding what to measure and how to share measurements are important considerations in all fields (Kenney, 2008; Langley et al., 2009). As will be seen, we found they also are critical in the study and improvement of preservice teacher education.

Openness and Measurement in Education

Although education is different from medicine and business in several important ways (e.g., the desired goals are more difficult to specify, the practices and desired outcomes are more difficult to measure), there are lessons for education to be drawn from the two principles. Openness could facilitate improvement by breaking down the boundaries between classrooms, an impediment to improvement noted years ago (Lortie, 1975). Openness has been shown to promote the collaboration among teachers frequently called for by educators studying the essential conditions for improving classroom teaching (Gallimore & Ermeling, 2010; Little, 1990; Vescio, Ross, & Adams, 2008).

Measurement of small tests of small changes that provide just enough data to make instructional decisions has been used in education for some time under the label "formative assessment" (Black & Wiliam, 1998; McManus, 2008; Wiliam, Lee, Harrison, & Black, 2004). Formative assessment can improve teachers' ability to make informed, data-based decisions. Although similar to formative assessment in some ways, the difference in the measurement process we report here is that the student learning data are used, not just to inform individual teachers' instructional decisions, but to inform the refinement of instructional products, products (like lesson plans) that hold knowledge for teaching and can be shared and passed along to new generations of instructors. The two principles—openness and measurement—were used by the first author to conduct two studies that investigated the conditions under which small amounts of data are used by instructors to make important changes to the course lessons.

The Context in Which Openness and Measurement Were Studied

Openness

Because of the culture that has been created in the mathematics K–8 teacher preparation program, a degree of openness already exists within the mathematics education group at our university. Multiple sections of three mathematics content courses are offered each semester. All instructors (faculty and doctoral students) of the same course use the same highly detailed lesson plans (see Hiebert & Morris (2009) and Morris (2012) for examples). Over time, instructors have become comfortable openly sharing their teaching experiences during the instructor weekly meetings. Because the same lesson plans are used, instructors are able to examine possible reasons for similar and different experiences. In the studies reported here, we were searching for the conditions of openness that would support the improvement of shared instructional products through the consistent (rather than the occasional) sharing and comparing of small amounts of data on preservice teachers' performance.

Measurement

Two kinds of measures have been used in our program to motivate and guide improvements to course lesson plans. As noted earlier, controlled experiments have been conducted, but these studies are too expensive in terms of time and effort to sustain a continuous improving system. Quiz and test scores are shared among instructors, but these measurements are too far removed from the daily lessons and not sufficiently detailed to prompt insights into problems that could be fixed or to create the moti-

vation needed for instructors to voluntarily go back and improve previous lessons. We have not yet identified the conditions that prompt instructors to consistently improve lessons based on systematically collected small amounts of lesson-relevant data.

Needed: Sharing Immediate Feedback of Preservice Teacher Performance

The following two studies describe an effort to measure preservice teachers' learning from individual lessons in a form that can be openly shared and discussed among the group of instructors to stimulate ideas for improving the lessons. With *clicker* technology now available, it is possible to create assessment items for this purpose.

Lesson-level clicker items offer two advantages: They are easy to administer, and they provide immediate feedback on how well preservice teachers in each section of a course understood a key aspect of the day's learning goal(s). This makes the data available for weekly meetings, where instructors focus on the most recently taught lessons. The items can be tightly connected to specific learning goals, making it easier for instructors to link preservice teachers' performance to particular instructional activities and teacher actions. These advantages could enable instructors to change the lesson based on preservice teachers' learning rather than on instructors' logic or anecdotes from a few vocal preservice teachers' responses.

What effect would using clicker items have on the nature of instructors' weekly discussions? Would data-based improvements be made to the lessons? The studies we describe in this article were designed to answer these questions. Because of space limitations, we will summarize the methods and results of Study 1 and present more details of Study 2.

Study 1: How Do Lesson-Level Assessment Data Affect Instructors' Efforts to Improve Lessons?

The overall question addressed in Study 1 was whether lesson-level data alone would drive the improvement process. Would simply having data about the effectiveness of lessons just taught affect the choices instructors make about how to improve the lessons? In particular, would the data prompt the instructors to engage in the kind of cause-effect reasoning that connects instructional moves to student outcomes (Gallimore, Ermeling, Saunders, & Goldenberg, 2009)? These general questions translated into three research questions:

1. Did the clicker data change the basis on which instructors revised lessons, from logic and anecdotes to student performance data?
2. What, exactly, did the instructors do with the data?
3. If instructors had access to the clicker data, were the changes they made to the lessons major (likely to affect student learning) or minor (less likely to affect student learning)?

As will be described, we found instructors could use the clicker data to develop hypotheses about major changes that might improve the lessons, but they did not do this frequently. Why?

Methods

All three instructors for the first mathematics content course for preservice K–8 teachers constituted the sample in Study 1. The course coordinator was a clinical faculty member (taught two sections); the other two instructors were doctoral students (taught one section each). As normal, the instructor group met weekly to discuss past and future lessons. At the beginning of the semester, the course coordinator sent the following instructions to the other two instructors, which set a typical agenda for the weekly meetings:

Before each meeting, please do the following so you are prepared for the discussion: (1) make a list of things that could be changed/improved for the 2 lessons we taught this week; (2) read through the 2 lessons to be taught next week; and (3) make a list of any questions you have about the lesson plans to be taught next week.

The first author wrote 25 clicker items (1–3 per lesson) to assess Lessons 7–19 (of 26).

Lessons 1–6 and Lessons 20–26 were not assessed with clicker items. This design provided an opportunity to compare the nature of the instructors' suggestions for improving the lessons before, while, and after using lesson-level clicker items. Items were given at the beginning of the lesson following the lesson that the items were assessing (e.g., Lesson 7 items were given at the beginning of Lesson 8). The clicker items were created to link tightly to the learning goals of each lesson. Here is an example.

Lesson 13 learning goal:

Given three numbers (e.g., 0.4, 6, 15) that can be related in a multiplication number sentence, preservice teachers will (a) arrange them into two valid equations (e.g., $0.4 \times 15 = ?$ and $15 \times 0.4 = ?$), (b) write the meaning of the equations (e.g., "find 4 groups of one tenth of 15," "find four one tenths of 15," "find four tenths of 15," "find 15 groups of 0.4"), (c) make [concrete] models of the equations, and (d) use the models and the meaning of the equations to help write a story situation that can be solved with the equation.

Clicker item for this goal:

A teacher asked her class to write a multiplication story problem that involved the numbers .7, 3, and 2.1 and could be solved with an equation of the form $a \times b = ?$. Four students wrote the following stories. Which of the stories are correct?

a. Brenda has 3 pounds of bananas. Sue has seven tenths more than Brenda. How many pounds of bananas does Sue have?

b. Bill has seven tenths of a pound of apples. Tom has 3 times more than Bill. How many pounds of apples does Tom have?

c. Howie has seven tenths of a pound of bananas. Stephen has 3 times as much as Howie. How many pounds of bananas does Stephen have?

d. Hilda has 3 pounds of apples. Jill has seven tenths as much as Hilda. How many pounds of apples does Jill have?

e. Both c and d are correct. (the correct answer)

The first author observed all instructor meetings, took detailed notes on the discussions of the instructors around the three agenda items, and collected the course coordinator's notes, which listed each of the proposed improvements. The first author has a collegial relationship with the instructors but did not supervise them.

Results and Discussion

Did the clicker data change the basis on which instructors revised lessons, from logic and anecdotes to student performance data? Instructors' discussions during the weekly meetings were coded to identify the rationales for the changes that were made. Three kinds of rationales were observed: (a) instructors' logic based on a rational analysis of the lesson, (b) anecdotes of individual students' responses during the lesson, and (c) clicker item data. For example, during the weekly meeting one instructor might say, "Sue asked me why we multiply to get the answer to this problem, so let's add an activity to make this clear." This was coded as an anecdote of individual students' comments. A second coder coded 20% of the data (the data for 15 of the 75 changes that the instructors decided to make over the course of the entire semester). Intercoder reliability was 93%.

On the first six lessons (not assessed with clicker items), the instructors suggested 16 changes to the lessons. The changes were motivated by 16 codeable rationales. The 16 rationales were about evenly split between instructor logic and anecdotes. On the next 13 lessons (which were assessed with clicker items), the instructors suggested 40 changes and formulated 43 rationales for those changes (a single change can be motivated by more than one rationale). Of the 43 rationales, 35 (81%) were based on instructor logic and anecdotes, and eight (19%) were based on the clicker data. On the final seven lessons (not assessed with clicker items), the instructors suggested 19 changes with 20 rationales. The basis for the rationales returned to pre-clicker patterns—the rationales were about evenly split between instructor logic and anecdotes. Instructors showed they could use the clicker data to make changes to the lessons, but they did so for only a small fraction of the lessons. What did they do with the data?

The answer to this second research question was found by coding what the instructors did with the data from each of the 25 clicker items into four categories that show increasing levels of analysis of the data: (1) no analysis, (2) discussed the meaning of the data (instructors tried to understand the nature of the student errors on the item), (3) connected the meaning of the data to their instruction (using their interpretation of student errors, instructors hypothesized instructional causes for these errors), and (4) made changes to the lesson based on the hypotheses. A second coder coded 40% of the data (the data for 10 of the 25 clicker items). Intercoder reliability was 97%.

Results showed that for 14 of the 25 items there was no analysis, for four items at least one instructor discussed the meaning of the data, for seven items the instructors worked together to connect the meaning of the data to hypotheses about a deficiency in the lesson that might explain the students' errors, and for these same seven items they made changes to the lesson based on these hypotheses. We will call these seven items *hypothesis items*. A hypothesis item is defined as an item that prompts instructors to formulate hypotheses about instructional causes for student performance on the item and to make changes to the lessons based on these hypotheses. Again, instructors showed they *could* engage in creating lesson

improvements based on a small set of relevant data, but they did not do this frequently.

Of the 25 items, the average performance across all sections for 13 items was below 70% and for 12 items the average performance was above 70%. All seven hypothesis items that prompted data-based changes were in the lower scoring set of items. In fact, 5 of the 7 hypothesis items were 5 of the 6 lowest scoring items. Apparently, instructors' attention was drawn to items on which the preservice teachers performed especially poorly.

Were the changes instructors made to the lessons major or minor? The changes instructors made during the clicker item lessons were coded into major changes (involved a significant change in the nature of an instructional activity, or the addition of a new activity) and minor changes (involved a small change in sequencing of activities, a slightly different explanation, or a small reduction in the number of problems). An example of a major change is adding a new activity to develop preservice teachers' understanding of the distributive property. Major changes are more likely than minor changes to affect preservice teachers' learning. For the major vs. minor code, a second coder coded 20% of the data. Reliability was 100%. Results suggested that basing changes on data led to major changes (5 of the 7 changes based on the seven hypothesis clicker items described above) whereas basing changes on instructor logic or anecdotes led to minor changes (30 of 33 changes based on instructor logic and anecdotes).

To summarize the results from Study 1, instructors showed they could use small amounts of data to make major changes to the lesson by forming hypotheses about what kind of change would improve preservice teachers' learning. But when just provided with the clicker items, they did not engage in this process very often. They did not connect the data to their instruction for 18 of the 25 items. When they did engage in this process, they selected lessons where the clicker data showed especially low performance. Moreover, for the 40 changes instructors made when clicker data were available, 32 of them were minor changes.

These results prompted a second study that tried to answer the following questions raised by these results. How can instructors be helped to engage in analyses of the clicker data more often, and even for items in which performance exceeds 70% correct? How can instructors be supported to use data analyses to engage in hypothesis-driven reasoning? How can instructors be encouraged to make major rather than minor changes to the lessons?

Because the results of Study 1 generated not only these questions but also some tentative answers, the second study was designed to address these questions by testing these answers. First, it seemed that instructors could be helped to engage in analyses of clicker data more often (first question posed above) if they were asked to share their clicker data for all items (including those showing high performance) during each weekly meeting. During Study 1, instructors chose to share clicker item scores for only 13 of the possible 100 scores (4 sections times 25 clicker items). In addition, to ensure that instructors evaluated all lessons, even those with clicker data above 70%, they could be asked, prior to each weekly meeting, to independently rate the effectiveness of each lesson after teaching it and decide whether improvements could be made. Continuous improvement requires constantly searching for ways to improve practices, even those that might appear to be *good enough*.

Second, we reasoned that engaging in data analyses could lead to more hypothesis-driven reasoning (second question above) if instructors could be supported in making their hypotheses explicit and making predictions about how preservice teachers' performance would improve on the items after lesson changes were implemented. During Study 1, when instructors engaged in data analysis, they looked at the poor performance of the preservice teachers, guessed what kind of change might improve this performance, changed the lesson accordingly, and left it to the Spring semester instructors to see if performance improved. Although they formed a hypothesis, they could not test their hypothesis. This makes it difficult to engage in the deeper cause-effect reasoning about instructional effects on learning that can produce *testable predictions*.

Finally, it appeared that major, rather than minor, changes could be facilitated (third question above) by pointing instructors to ideas or principles that could guide their search for substantive instructional changes that would likely make a difference for students' learning. Instructors could, for example, be asked to evaluate every lesson for whether the nature of the learning opportunities was consistent with the learning principles that drove the creation of the lessons.

Study 2: What Features Facilitate a Data-Based Sustaining Improvement System?

Guided by the questions raised by Study 1 (conducted during the Fall semester) and our tentative explanations or answers to these questions, Study 2 (conducted during the following Spring semester) was designed to change the conditions from Study 1 to test these explanations. Three

requests of the instructors changed the conditions from Study 1. First, instructors were asked to use the first few weekly meetings of the semester to look back at the clicker data from the previous semester, compare data across sections for each item, and make changes to the lessons when their analyses of the data suggested changes were needed. This was called the *Course Improvement Project*. The instructors were given an *Activity Change Report* form (**Appendix A** shows the form and the instructor's reports) to carry out the project. The aim was to engage instructors in using the clicker data from the previous semester in hopes of stimulating the formation of data-based hypotheses about changes to the lessons that could be tested during the current semester, thereby enabling the formation of predictions. Comparing Fall data to Spring data would allow instructors to create *forward-looking* hypotheses—hypotheses that included predictions that instructors could test during the semester.

The second request asked instructors to individually fill out an *Instructor Booklet* in which they entered the percentages of students in their section(s) who gave each multiple-choice response to the clicker item and then wrote a response to the following questions: "Based on the data, do you think there is a problem with the lesson that needs to be fixed? If you responded 'yes,' do you have a suggestion about how it can be fixed?" The aim of the first question was to encourage instructors to think about all lessons, even those in which performance was above 70%. The aim of the second question was to encourage instructors to formulate cause-effect hypotheses that connected the clicker data to their instruction. The booklets also contained Effectiveness Score sheets. These sheets asked instructors to give each taught lesson an Effectiveness Score (see **Appendix B**). The Effectiveness Score was the sum of the instructor's ratings on the overall effectiveness of the lesson plus the lesson's effectiveness in engaging the preservice teachers in dimensions of the two learning principles underlying the lesson design: opportunities for preservice teachers to (1) engage in productive struggle with the important mathematical ideas and (2) to construct explicit understandings of the key conceptual relationships that would support students' efforts to achieve the learning goal. The focus on the two pedagogical principles was intended to remind instructors of where they could find ideas that would produce major, rather than minor, changes to the lesson.

The third request asked instructors to share the clicker data and the Effectiveness Scores during the weekly meetings. The process for sharing the data was developed by the course coordinator: She wrote on the board the percentages of students who had correctly responded to each of the clicker items administered that week. She provided this information for seven sections individually—the four Fall sections (from Study 1) and the three Spring sections.

Methods

Sample. Two instructors who had participated in Study 1, one of the doctoral students and the clinical faculty member, participated in Study 2 and taught the three sections offered that semester. The clinical faculty member served as the course coordinator and taught two sections of the course; the doctoral student taught one section.

Methods. The 25 clicker items from Fall were used in Spring, along with one new clicker item written by the course coordinator. The first author observed all of the weekly instructor meetings and took notes on instructors' discussions of the following: clicker data, Effectiveness Scores, proposed revisions to the instruction, rationales for revisions, and the Course Improvement Project. Additional sources of data included the course coordinator's notes, the Instructor Booklets, and the Activity Change Reports from the Course Improvement Project (Appendix A). The same codes used in Study 1 were used in Study 2.

Results and Discussion

Did new conditions promote data-based improvements of lessons? An initial question is whether the change in conditions prompted a change in the way the clicker item data were used to stimulate changes to the lessons. The simple answer is yes, but with some critical caveats.

Formulating hypotheses and testing predictions increased instructors' use of data. As in Study 1, only Lessons 7–19 were assessed using clicker items. For changes suggested to these 13 lessons, 38% of the instructors' rationales for making the changes were based on clicker item results. Even for changes to the first six lessons of the course (which were not assessed with clicker items), 27% of the rationales appealed to students' performance on clicker items (while teaching lessons 7–19, instructors went back and made changes to the first six lessons based on the clicker results to these later lessons). Recall that corresponding percentages for Study 1 were 19% and 0%, respectively.

The most plausible explanation for instructors' increased use of data in Study 2 comes from the effects of the Course Improvement Project. Recall that during the first weeks of the semester, instructors were asked to review the Fall semester clicker data. As they looked across the Fall sections, they noticed a pattern of low performance in some lessons. They concluded the clicker items of interest all required a competency not addressed in the course (i.e.,

preservice teachers did not “flexibly interpret diagrams, including K–8 students’ diagrams”). The instructors designed several instructional activities to develop the competency, implemented them in the Spring, and compared the Fall and Spring data to test several hypotheses (completed in several cycles described below): The new instructional activities would develop the competency, the set of clicker items all required the competency and performance on all the items would therefore improve, and the competency would transfer to related tasks and clicker items. The instructors then used the identified set of clicker items to test the changes they had made to accomplish their goal for improvement. At this point in the semester, the only request that had been made of the instructors was to fill out the Activity Change Report form.

Generating these hypotheses about what might have caused preservice teachers’ low performance shifted the instructors’ reasoning to a forward direction; they now were making changes to lessons, predicting their effects, and using and comparing the Fall data (collected before the changes were implemented) and the Spring data (collected after the changes were implemented) to test their predictions. This concept of forward reasoning now allows us to contrast the kind of reasoning instructors were faced with in Study 1: reasoning backward from the clicker data about what might have caused difficulties for preservice teachers. We will call this analysis *single-item retrospective analysis*—wondering what went wrong in an individual lesson without the benefit of any predictions about the outcomes. It is difficult to interpret data if explanatory hypotheses and predictions have not been made prior to collecting the data. It still is possible to formulate hypotheses based on retrospective analysis, but the hypotheses cannot be immediately tested. We propose that forward-looking hypothesis formulation and the testable predictions it generates motivates increased use of performance data.

Cycles of hypothesis testing prompted unpacking the goal for improvement. Once the instructors generated hypotheses about what changes to instruction might facilitate preservice teachers’ flexibly interpreting diagrams, they carried out their hypothesis formulation, predictions, and testing in three cycles. Each cycle involved hypothesizing and making a change in the lessons, predicting the effects of the change, implementing the change and testing the effects by comparing the Fall data with the Spring data as the Spring data became available, studying the results, and revising for another cycle. The first cycle was conducted in the context of addition and subtraction, the second in the context of multiplication, and the third in the context of division. In each cycle, instructors were testing hypotheses about how their revised lessons might help preservice teachers interpret diagrams more flexibly.

Through engaging in cycles of hypothesis testing, the instructors identified several factors that appeared to affect students’ ability to flexibly interpret diagrams: understanding that some number sentences can be interpreted and modeled in multiple ways, understanding the distributive property, understanding that diagrams are ambiguous if the units of measure are not explicitly identified, and understanding that quantities need to be labeled. This analysis prompted them to make further changes to the lessons after each of the first two cycles, the cycle focused on addition and subtraction and the one focused on multiplication.

By the third cycle, the instructors hypothesized that increased flexibility with diagrams would transfer to the operation of division without adding new instructional activities specifically about division. They compared the Fall and Spring data for clicker items 23 and 26 to test this hypothesis. The following discussion at an instructor meeting after the third cycle shows their subsequent attempts to *unpack the competencies* that affect students’ ability to flexibly interpret diagrams in the context of division, to refine the clicker items based on this analysis, and, consequently, to understand students’ thinking and difficulties at a deeper level.

Instructor A: Clicker question 23 (see Appendix A) showed [the preservice teachers] were not able to transfer some of the ideas about diagramming that we discussed for addition, subtraction, and multiplication [through the new assignments]—that is, being flexible about what the diagram represents. In Question 23, the diagram could represent either meaning of division [repeated subtraction or partitioning] and they didn’t see that. They tended to see partitioning, but it could also be representing repeated subtraction. They might not be strong enough with both meanings of division so they only see one or the other. Being flexible with “what does this diagram represent?” did not transfer, but I think it’s more that they aren’t strong enough with division.

Instructor B: This clicker item is more about the meaning of division. They’re still muddling the two interpretations together. Maybe it’s too early to test for transfer.

Instructor A: So instead, maybe we could have a clicker item [at this point in the course] that tests their ability to transfer to division the idea of not having measuring units provided.

The choices would not be focused on the meanings of division, but would ask them "which number sentences could this diagram represent?" This would test their ability to transfer the idea that we have to label the measuring units or the diagram is ambiguous. Then a little later, we would have an item where the options focus on the two meanings of division, when they have a stronger understanding of the two meanings of division. Also, the students need to know that when there's not enough information, you need to be more flexible. I looked at Exam 2 problem 1 for another example of student diagrams involving division to see if they do better when it's more constrained; when they get more information on the diagram, do they do better? From the problem, I found they do okay when more information is included.

Instructor B: So maybe they don't do as well when diagrams are missing information and when they can be interpreted in multiple ways.

This conversation shows a deepening analysis of the learning goal—the instructors' target for improvement—into constituent parts. At the beginning of the third cycle, they expected "flexibility" to transfer to the operation of division. Based on multiple cycles of testing, they now hypothesized that the ability to flexibly interpret diagrams requires some concepts that cut across the four operations (e.g., the necessity of specifying units) and some concepts specific to the operation (e.g., the two meanings of division), and flexible interpretation can only be expected after those subconcepts are sufficiently robust. The instructors began to design clicker measures that would directly test their hypotheses about these specific competencies. The items would assess just one skill, and the data would then have more precise implications for instruction.

In an effort to gather even more data to inform their analysis of preservice teachers' thinking about these issues, the instructors, without prompting from the authors, looked at responses to particular exam questions. Within the context of testing their increasingly refined and forward-looking hypotheses, these exam questions no longer seemed distant and unhelpful for improvement but rather added useful information to the instructors' hypotheses for improvement. An important parenthetical point here is that the analysis of exam responses, within the context of cause-effect hypotheses of teaching-learning, is an activity that can be engaged by all teacher educators (and classroom teachers). Special clicker items are not required.

Stepping back, we believe a key lesson to draw from the cycles of hypothesis testing is that the decomposition of an initial, often vaguely defined learning goal into its component parts is a critical outcome of hypothesis formulation and refinement during sequential cycles of testing.

Unless teachers can unpack often generally stated learning goals into more precise component parts, they cannot design appropriate instruction, they cannot focus their improvement efforts in productive ways, and they cannot measure precisely enough students' performance to know if their changes are actual improvements. Unpacking learning goals is especially difficult for teachers (Morris, Hiebert, & Spitzer, 2009). Consequently, the fact that instructors in this study engaged in this unpacking process as they were driven to test their hypotheses through repeated cycles of testing instructional changes suggests this as a productive setting in which teacher educators, and classroom teachers, could decompose learning goals in a useful way.

Hypothesis testing reduced instructors' tolerance for poor performance. If instructors actually tolerate students' performance until it reaches some low threshold level, this obviously would interfere with efforts to constantly improve all the lessons. The first question in the Instructor Booklets measured this tolerance (recall that the Instructor Booklets asked instructors to individually enter the clicker data for each lesson for their section(s) and then decide whether the lesson needed improvement). As found in Study 1, when just looking at performance data, instructors in Study 2 indicated that lessons did not need improvement until clicker item performance fell below 70%. The question is: How can instructors motivate themselves to improve lessons in which the performance data rises above 70%? The answer suggested by Study 2 is that instructors should formulate forward-looking hypotheses about how to improve the lessons and engage in cycles of testing.

Evidence for the claim that hypothesis formulation and testing reduced instructors' tolerance for marginal performance can be found in the instructors' report on their Course Improvement Project (Appendix A). After their first cycle of testing focused on addition and subtraction, the instructors reported the results for clicker item 10, an item used to test the effects of a new instructional activity. The percentages of correct responses for the three sections in the Spring (after their intervention was implemented) were 82%, 76%, and 87%. Although all of the percentages exceeded the instructors' prior 70% cut-off for tolerating students' marginal performance, they pushed for more improvements in their Activity Change Report written near the end of the semester (see italicized portion of Appendix A, item 9 under Question 10).

More systematically analyzed data that support the claim that hypothesis testing reduced instructors' tolerance for marginal performance are found in Table 1. Table 1 includes the Spring data for all clicker items, divided into hypothesis and nonhypothesis items. Recall that hypothesis items are items that prompted instructors to make hypotheses about a deficiency in the lesson that might explain the students' errors and to make changes to the lesson based on these hypotheses. By comparing the Fall and Spring data, the nine hypothesis items shown in Table 1 measured the effects of the changes made in the Fall (Study 1) or in the Spring Course Improvement Project. *Nonhypothesis items* are clicker items not used to test hypotheses.

Data in the first column of the table show the preservice teachers' results for each clicker item. Data in the second column are from the Instructor Booklets. They are the instructors' individual responses to the question "Based on the data [their students' performance on the specific clicker item], do you think there is a problem with the lesson that needs to be fixed?" If the entry in the second column says "no," it means both instructors independently responded "no" to the question. If the entry says "yes," it means both instructors independently responded "yes" to the question. If the entry says "yes; no," it means Instructor A responded "yes" and Instructor B responded "no." To detect patterns, items are ordered, within nonhypothesis and hypothesis categories, with respect to instructors' belief that a lesson needed to be fixed. The shaded rows in the table highlight the clicker items that prompted at least one instructor to believe a lesson needed to be fixed. Comparing columns 1 and 2 provides a measure of the instructors' tolerance for marginal performance. Notice that when they were assessing the effectiveness of a lesson on their own, their tolerance was similar to that in Study 1—they did not believe lessons needed improving if students' performance in their section(s) was greater than 70% correct. This is apparent by the fact that, with one exception, items with at least one score below 70% are exactly those items in the shaded portions of the table.

Columns 3, 4, and 5 of Table 1 show how the instructors analyzed the data for each item during the weekly meetings as the Spring data became available and they could compare Fall and Spring data. The Spring data allowed the instructors to test their hypotheses about changes they had made to the lessons, both those made in the Fall and those made in the Spring Course Improvement Project (would the change improve student performance?). Changes to the lessons shown in column 5 are changes instructors made in the Spring after they had compared Fall and Spring data.

For the 12 nonhypothesis items where both instructors independently indicated in their Instructor Booklets that the lesson did not need to be fixed, the instructors analyzed the meaning of students' wrong answers for only 3 of the 12 items (column 3) and made no changes to the lessons based on the Spring data (column 5). For nonhypothesis items where one or both instructors independently indicated that the lesson needed to be fixed because of poor student results, the instructors analyzed the meaning of students' wrong answers at a slightly higher rate (for 2 of 5 items versus 3 of 12 items) but were not likely to change the lessons (this occurred for only 1 of the 5 items and the revision was a minor change). These results are similar to Study 1 and show that instructors often tolerate their students' performance *if* it rises above 70% and no hypotheses are driving improvement efforts.

As shown in columns 3, 4, and 5 of Table 1, treatment of the hypothesis items was different. Discussions of hypothesis items usually included analyzing the meaning of students' wrong answers *and* explicitly connecting the data to the instruction (7 of 9 hypothesis items versus 3 of 17 nonhypothesis items). The reason for the high rate for hypothesis items might be obvious but it is critical: Unlike retrospective analysis, the instructors' hypotheses identify an instructional cause and they interpret student responses within that framework. Rather than moving backward from an effect (the student data) to the identification of a cause (something in their instruction) as they must do in retrospective analysis, the hypothesis stipulates the cause (the changes they made to the instruction), and instructors use the data to test whether their hypotheses are correct. As noted earlier, this forward-looking analysis appears to encourage the use of data.

Table 1
Instructors' Analyses of the Spring Clicker Data for Hypothesis and Nonhypothesis Items

Percentage of correct responses on each clicker item in each of the three sections (Item number)	From instructor booklet: Based on the clicker data from their sections, did the instructors believe the lesson needed to be fixed?	Spring weekly meetings: Did instructors analyze the meaning of students' wrong answers?	Spring weekly meetings: Did instructors connect the data to instruction?	Spring weekly meetings: Did instructors decide to make additional changes to the lesson after they analyzed the data? If so, was it a major or minor change?
Nonhypothesis Clicker Items				
95, 97, 94 (Item 7)	No	No	No	No
91, 97, 94 (Item 14)	No	No	No	No
90, 88, 89 (Item 11)	No	No	No	No
95, 86, 83 (Item 12)	No	No	No	No
78, 89, 94 (Item 15)	No	No	No	No
81, 82, 94 (Item 8)	No	No	No	No
81, 76, 89 (Item 5)	No	No	No	No
74, 85, 72 (Item 1)	No	No	No	No
68, 73, 79 (Item 3)	No	No	No	No
74, 73, 76 (Item 24)	No	Yes	No	No
100, 100, 100 (Item 22)	No	Yes	Yes	No
91, 87, 88 (Item 2)	No	Yes	Yes	No
83, 53, 94 (Item 13)	Yes; no	Yes	Yes	Minor
48, 87, 82 (Item 17)	Yes; no	Yes	No	No
62, 64, 56 (Item 4)	Yes	No	No	No
76, 58, 56 (Item 6)	Yes	No	No	No
64, 59, 56 (Item 9)	Yes	No	No	No

Table 1—Continued

Percentage of correct responses on each clicker item in each of the three sections (Item number)	From instructor booklet: Based on the clicker data from their sections, did the instructors believe the lesson needed to be fixed?	Spring weekly meetings: Did instructors analyze the meaning of students' wrong answers?	Spring weekly meetings: Did instructors connect the data to instruction?	Spring weekly meetings: Did instructors decide to make additional changes to the lesson after they analyzed the data? If so, was it a major or minor change?
Hypothesis Clicker Items				
77, 73, 100 (Item 19)	No	No	No	No
77, 76, 94 (Item 20)	No	No	No	No
76, 87, 82 (Item 10)	No	Yes	Yes	No
86, 85, 100 (Item 18)	No	Yes	Yes	No
42, 75, 73 (Item 26)	Yes; no	Yes	Yes	Major
41, 67, 59 (Item 16)	Yes	Yes	Yes	Major
48, 46, 59 (Item 21)	Yes	Yes	Yes	Major
43, 61, 61 (Item 23)	Yes	Yes	Yes	Major
56, 60, 61 (Item 25)	Yes	Yes	Yes	Major

There was also evidence that the instructors were less tolerant of marginal performance in the context of lower-scoring hypothesis items versus lower-scoring nonhypothesis items (items in the shaded portions of Table 1). These two sets of items were treated very differently. Instructors analyzed the meaning of students' wrong answers and connected the data to the instruction (columns 3 and 4) for all five lower-scoring hypothesis items but carried out this kind of analysis for only one of the five lower-scoring nonhypothesis items. The analyses of the five lower-scoring hypothesis items all resulted in major revisions as opposed to one minor revision for the five lower-scoring nonhypothesis items (column 5).

To summarize, the results of Study 2 suggest that the kinds of data comparisons that motivate data-based improvement efforts, regardless of performance levels, are comparisons of performance before-instructional-change to after-instructional-change involving a hypothesis test. Forward-looking hypothesis testing seemed to motivate improvement efforts and decrease instructors' tolerance for marginal performance.

Non-pooled data comparing instructors' results did not promote data-based improvements. Recall that a goal of these studies was to search for conditions of open sharing of data that promote continuous improvement. A common approach to motivating U.S. teachers to improve their performance is to compare teachers in terms of their students' performance (Cohen, 1996). Indeed, debates currently are raging about how to conduct such comparisons in equitable ways (e.g., Harris, 2009). The important question for us was whether the conventional method of comparing instructors in terms of their students' achievement worked for jointly improving instructional products. How did instructors respond to the course coordinator posting clicker results, by section, for the Spring semester, at the beginning of each weekly meeting?

One way to answer the question is to check what happened when the performance differences were large, say about 25 percentage points or more. If comparing data across instructors is a condition that motivates them to examine the data and search for improvements, these big differences in performance should prompt discussions among instructors that would explore the reasons for the better performance. There were five such items. For one of the items (Item 19), Instructor B's section scored 23% and 27% higher than Instructor A's two sections respectively. But the instructors carried out no analysis of the item. Instructor A did not ask Instructor B what about her instruction might have produced the higher performance.

For the remaining four items, one of Instructor A's sections (not always the same section) scored noticeably lower than the other two sections. When the data were shared for two of the four items (Items 16 and 26), the instructors did not refer to the performance differences across sections and the data did not prompt them to compare their instruction in the three sections. For the remaining two items (Items 13 and 17), the instructors made some attempt to explain away the data; they attributed the results to some aspect of the students' behavior (e.g., lack of studying) rather than the instruction. For example, for Item 17, the following exchange took place.

> *Instructor A:* [Puts the data on the board: 82% correct (Instructor B's section), 48% correct (Instructor A's first section), 87% correct (Instructor A's second section)]
>
> *Instructor B:* [Laughing] What's with the section? Maybe you do better the second time around?
>
> *Instructor A:* Thirty-three percent said D was the answer. I asked, "Why is B not correct?" They just said, "Oh. They weren't careful."

Thus for all five items showing a large performance difference, the open sharing of data that compared the performance of the instructors' students did not prompt the instructors to compare the details of their instruction in order to identify potentially better practices.

Because comparing student performance data across instructors is often assumed to create conditions that motivate instructors to improve, we consider why this did not happen, even for these large differences. When instructors are implementing the same lessons using similar instructional practices, student performance measures can show one instructor or class is doing better, but the instructors cannot explain why. If the differences were caused by variants of the intended instruction, the instructors could not reconstruct these details and consequently had no basis on which to hypothesize causes for the differences. So, in some cases they did what many instructors do—they attributed the results to students' behavior. This is a problem for a continuous improvement process because it means that instructors, at least momentarily, are relinquishing responsibility for the effects of their instruction and removing themselves from the improvement process (Kenney, 2008).

A second problem with the instructors' comparing section-by-section data (four sections from the Fall and three sections from the Spring) was that the data were too difficult to interpret in this form; the instructors could not always tell whether the instructional changes they were making in the Spring were improvements. For example, in the instructors' report of the Course Improvement Project (Appendix A), the instructors compared the percentages of correct responses in the seven sections and made the following statement.

> **For clicker items 23 and 26, the instructors thought that the data showed slight improvement, but we were not satisfied. We thought that [preservice teachers] would be able to transfer the types of knowledge we discussed with interpreting diagrams for addition, subtraction, and multiplication to division diagrams. However, the data showed that this is not really the case.**

The results from the two items were treated the same ("the data showed that this is not really the case [that transfer would occur]"). However, if the data from all classes in a given semester are pooled, a comparison reveals students performed significantly better in the Spring (after the instructional changes were implemented) than they did in the Fall (before the changes were implemented) on item 23 ($z = 3.47$, $p < .001$) whereas there were no significant differences between the Fall and the Spring for item 26. This type of comparison would have provided information the instructors needed when they were attempting to unpack the competencies that contribute to students' ability to flexibly interpret diagrams (described above). They could have analyzed the competencies that are involved in answering item 23 versus item 26 and used that analysis to understand the effects of their new instruction.

These two issues indicate that the section-by-section sharing of data was problematic. What kinds of comparisons would be more productive? We can offer the following suggestions.

Results from Study 2 suggest that before- and after-instructional-change comparisons involving a hypothesis test changed the acceptable level of failure and motivated data-based improvement efforts. So, comparisons that support hypothesis testing appear to be more productive. The findings above suggest this happens when data are compared across time but not across instructors or classrooms. On a practical level, this means pooling the data across classes for each clicker item and comparing the data across semesters.

Comparing data across time supports hypothesis testing *when* instructors can distinguish improvements from just changes. One way to enable these distinctions is to statistically analyze the pooled data across semesters. These analyses can determine whether a change to the instruction was a real improvement. Because some changes produce important but small improvements, several hypothesis-testing cycles might be needed to accomplish an improvement goal. A test of significance can be used to compare the beginning overall percentage of correct responses (before any changes are made) with the final overall percentage of correct responses (after a number of related revisions have been made to address the same goal).

A consequence of statistically testing pooled performance data across time is that it places attention where it belongs—on the instructional products (e.g., lesson plans) rather than on individual instructors. A shared instructional product is designed to minimize across-instructor variability so the instruction can be linked to student outcomes (Morris & Hiebert, 2011). Comparing data across instructors highlights this variability and focuses attention back onto the individual instructors. A before- and after-instructional-change comparison, with data pooled across instructors, returns the focus to the instructional product.

Before concluding this section, we need to add an important caveat. Past experience has taught us that, under some conditions, comparing results across instructors can be productive. If instructors are discussing relative successes and failures of recent lessons and find that one instructor had more success than others, this discussion can encourage the successful instructor to share her approach with other instructors often saying, “Oh, that’s a good idea. I’ll try that,” usually to good effect. The difference in the studies we are reporting here is that the clicker *data* created differences among instructors in preservice teachers’ performance with no hypotheses to explain the differences. This left the instructors with no productive avenue of analysis.

What effects did the Effectiveness Scores have on the instructors’ improvement efforts? Effectiveness Scores had two effects on improvement efforts. First, they drew attention to instructional problems. The instructors were more likely to think there were problems with the instruction that needed to be fixed when they were assigning Effectiveness Scores than when they were evaluating the clicker data. Instructor A and Instructor B believed that 38% and 27%, respectively, of the lessons with clicker items needed to be fixed, based on the clicker data. In contrast, Instructors A and Instructor B gave Effectiveness Scores indicating there was some deficiency in the lesson to 69% and 54%, respectively, of the lessons assessed by clicker items.

Second, the Effectiveness Scores prompted major instructional changes rather than just minor changes. This was most pronounced for the productive student struggle dimensions of the Effectiveness Score. Instructor A and Instructor B assigned non-perfect scores to one or both of the student struggle dimensions for 33% and 37% of all course lessons respectively. Productive student struggle then became a major focus of the discussions during the instructor meetings. Whereas two changes in the Fall semester involved increasing productive student struggle (both major changes), 13 changes of this type were made in the Spring (11 were major changes). In contrast to the tendency of the instructors in the Fall to make minor changes (see Study 1), the instructors in the Spring made a number of major changes by developing cognitively demanding activities to increase the level of productive student struggle.

Based on these two results, it appears that an especially useful role for Effectiveness Scores is to point instructors to fundamental features of lessons that could be considered for improvement. On a daily basis, as instructors assign each lesson a score, they are reminded of learning principles that drive the lessons and could become the target for cycles of specific testable major changes.

In this study, the Effectiveness Scores focused the instructors’ attention on productive student struggle. Typical of fundamental learning goals in education, productive struggle is quite vague. Without a more refined and elaborated definition, how do teachers improve lessons with respect to this learning principle? The results of Study 2 reveal that one significant contribution of the repeated cycles of hypothesis testing can be clarifying the meaning of these principles and enabling more focused improvement efforts.

In future semesters, other foundational principles could be selected. In this way, teacher educators can customize Effectiveness Scores to shine the spotlight on learning principles on which they would like to focus their improvement process.

Does collecting and providing lesson-level data to instructors produce increased learning? The ultimate question is whether all of this work to improve the quality of the instructional products increases preservice teachers' learning. To answer this question, clicker data from all sections in the Fall and the Spring, respectively, were pooled for each of the 25 items that were administered in both Fall and Spring and the percentages of correct responses were compared across semesters. The comparisons test the effects of the lesson changes that were made before administering the Spring clicker items. Z-scores were calculated and a 99% confidence level was used as the standard for increased learning. Percentages of correct responses did not significantly improve for any of the 16 nonhypothesis items (see Table 1 for these items). For the hypothesis items, performance increased significantly for 5 of the 9 items (Items 10, 16, 18, 21, and 23). For one of the remaining four items (Item 25) only minor changes were made and consequently performance increases would not be as likely. For the final three items (Items 19, 20, and 26), major changes were made but no significant performance increases were found on these items. This might be due to the difficulty of the mathematical topics being studied; several cycles of improvement work might be needed.

Implications for Teacher Education

The promise of the principles of openness and measurement to make productive use of small amounts of data was realized, but it was the details of applications that mattered. Given increased attention to data-driven instruction, in both K–12 classrooms and teacher education programs, the details become critical. Taken together, Studies 1 and 2 suggest that just providing instructors data on students' performance is not enough to prompt frequent and continuing improvements in instruction. Data on past performance, even at the lesson level, do not, by themselves suggest what changes should be made. Perhaps because the data do not reveal immediate implications for instruction, they do not seem to motivate instructors to engage in the time-consuming work of analyzing instruction and learning in a cause-effect way and making improvements.

What seems to be critical is to provide instructors (who are teaching toward the same learning goals) with guidelines of how to *use* the performance data. In particular, the results of these studies suggest that productive work is facilitated by asking instructors to use small amounts of targeted performance data to formulate hypotheses about what changes to instruction might improve performance, to make these changes and predict what effect they will have, and then to measure the effects of these changes. The results can then prompt a new cycle of hypothesis development, changes to instruction with predicted effects, and measurement. These hypothesis-testing cycles seemed to be the key to instructors' success in making major improvements to the lessons of interest.

One consequence of these testing cycles appeared to be a decrease in instructors' tolerance of preservice teachers' marginal performance. Apparently, instructors' views can change from "that performance is good enough" to the view that marginal performance is something that can be studied further and improved even more ("Try again. Fail again. Fail better" (Beckett, 1983).).

A second consequence of hypothesis testing cycles appeared to be a clarification of the learning goals into more specific subgoals or parts that can be studied with more precision. A somewhat unique feature of education compared with other fields is that many learning goals are quite general. Helping students make progress in achieving them requires breaking them into constituent parts and studying how to help students achieve more precisely stated subgoals.

We began by borrowing from traditions of improvement in other fields, especially medicine and business, to hypothesize that the principles of openness and measurement could play an important role in improving teacher education, and teaching more generally. We now look back to this tradition of improvement and note that the cycles of hypothesis testing we described bear a striking resemblance to something called PDSA (Plan, Do, Study, Adapt) cycles in Improvement Science (Langley et al., 2009; Nolan, Schall, Berwick, & Roessner, 1996).

The PDSA cycle is an iterative process in which improvements are developed, tried, studied, and then refined, over and over, until quality is improved. The first step is to *plan* the test. The plan involves creating a potential solution based on a clear hypothesis that generates an explicit, testable prediction about the outcomes. The second step is to *do* or carry out the test—that is, to execute the plan and assess the results of it by collecting small amounts of relevant data. Next comes *studying* the result and comparing the result to the prediction. It is this comparison that produces learning. Finally, based on the results of the analysis, a decision is made on how to act. What next step is warranted based on what was learned? Should the change be adopted, adapted and tested further, or abandoned? We believe that the improvement process in teaching could benefit from learning more about the processes that Improvement Science has to offer.

Finally, we believe the results indicate that the *method* of sharing and comparing data matters. Many educators have too quickly accepted the assumption that comparing teachers on their students' performance will lead to improved teaching. As we have described, other methods

of sharing and comparing these data might be more productive. For example, comparing data across time seems to work better than comparing data across instructors.

Our aim in these two studies was to learn something about the conditions under which the open sharing of small amounts of data can, and will, be used by instructors to improve their instructional products. We hope the suggestions we offered will be tested further, both in teacher education and classroom settings.

References

Austin, J. (2012). *Continuous improvement of mathematics teacher education.* Presentation at the annual meeting of the National Council of Teachers of Mathematics. Philadelphia, PA.

Beckett, S. (1983). *Worstward ho*. New York: Grove Press.

Berk, D., & Hiebert, J. (2009). Improving the mathematics preparation of elementary teachers one lesson at a time. *Teachers and Teaching—Theory and Practice*, *15*(3), 337–356.

Berk, D., Taber, S. B., Gorowara, C. C., & Poetzl, C. (2009). Developing prospective elementary teachers' flexibility in the domain of proportional reasoning. *Mathematical Thinking and Learning*, *11*(3), 113–135.

Berwick, D. M. (2003). Errors today and errors tomorrow. *New England Journal of Medicin*e, *348*(25), 2570–2572.

Berwick, D. M. (2005, December 12). Keys to safer hospitals: A set of simple precautions could prevent 100,000 needless deaths every year. *Newsweek*, *146*(24), 76–78.

Berwick, D. M. (2008). The science of improvement. *Journal of the American Medical Association*, *299*(10), 1182–1184.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, *5*(1), 7–74.

Cohen, D. K. (1996). Rewarding teachers for student performance. In S. H. Fuhrman & J. A. O'Day (Eds.), *Reward and reform* (pp. 60–112). San Francisco, CA: Jossey-Bass.

Gallimore, R., & Ermeling, B. A. (2010). Five keys to effective teacher learning teams. *Education Week*, *29*(29).

Gallimore, R., Ermeling, B. A., Saunders, W. M., & Goldenberg, C. (2009). Moving the learning of teaching closer to practice: Teacher education implications of school-based inquiry teams. *Elementary School Journal*, *109*, 537–553.

Gawande, A. (2007). *Better: A surgeon's notes on performance.* New York, NY: Metropolitan Books/ Henry Holt.

Harris, D. N. (2009). Would accountability based on teacher value added be smart policy? An examination of the statistical properties and policy alternatives. *Education Finance and Policy*, *4*, 319–350.

Hiebert, J., & Morris, A. K. (2009). Building a knowledge base for teacher education: An experience in K–8 mathematics teacher education. *Elementary School Journal*, *109*(5), 475–490.

Kenney, C. (2008). *The best practice: How the new quality movement is transforming medicine.* New York, NY: PublicAffairs.

Langley, G. J., Moen, R., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: A practical approach to enhancing organizational performance* (2nd ed). San Francisco, CA: Wiley.

Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record*, *91*, 509–536.

Lortie, D. (1975). *Schoolteacher: A Sociological Study.* Chicago, IL: University of Chicago Press.

McManus, S. (2008). *Attributes of effective formative assessment.* Washington, DC: Council of Chief State School Officers.

Morris, A. K. (2012). Using "lack of fidelity" to improve teaching. *Mathematics Teacher Educator*, *1*(1), 71–101.

Morris, A. K., & Hiebert, J. (2011). Creating shared instructional products: An alternative approach to improving teaching. *Educational Researcher*, *40*(1), 5–14.

Morris, A. K., Hiebert, J., & Spitzer, S. M. (2009). Mathematical knowledge for teaching in planning and evaluating instruction: What can preservice teachers learn? *Journal for Research in Mathematics Education*, *40*(5), 491–529.

Nolan, T. W., Schall, M. W., Berwick, D. M., & Roessner, J. (1996). *Reducing delays and waiting times throughout the healthcare system*. Boston, MA: IHI.

Rother, M. (2009). *Toyota Kata: Managing people for improvement, adaptiveness, and superior results.* New York, NY: McGraw-Hill.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, *24*, 80–91.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, *11*(1), 49–65.

Authors

Anne K. Morris, School of Education, 105B Willard Hall, University of Delaware, Newark DE 19716; abmorris@udel.edu

James Hiebert, School of Education, 107A Willard Hall,
University of Delaware, Newark DE 19716;
hiebert@udel.edu

Appendix A: Study 2: Activity Change Report Forms That Were Completed by the Instructors to Report on Their Course Improvement Project

Activity Change Report

An instructor group should fill out this form whenever they change an activity in the lesson plans in response to the student results from a clicker item. **Please fill out one form for each activity that is changed.** In some cases, a clicker item or items might suggest several activities should be changed. For example, students' performance on a number of items might suggest that the lessons are not effective in developing pre-service teachers' ability to write story problems. The instructor group might then develop a rationale (see #5 in the form) to address this issue. For example, they might identify a new way of teaching the writing of story problems that they believe will be more effective. This rationale—that pre-service teachers are not performing well on story problem clicker items and that this new approach to the teaching of writing story problems will be more effective—has implications for several lessons (e.g., lessons on writing story problems for addition and subtraction, lessons on writing story problems for multiplication and division). The instructor group might then change several activities across several lessons to address this rationale. If multiple activities are changed because of the instructor group's rationale, then the separate forms for each activity should be stapled together. This will allow subsequent instructor groups to see all of the changes that were made to address that rationale. In other words, if a common rationale underlies a group of changes, please staple the separate forms together; activity change report forms should be grouped by rationale.

Instructors' Activity Change Report I

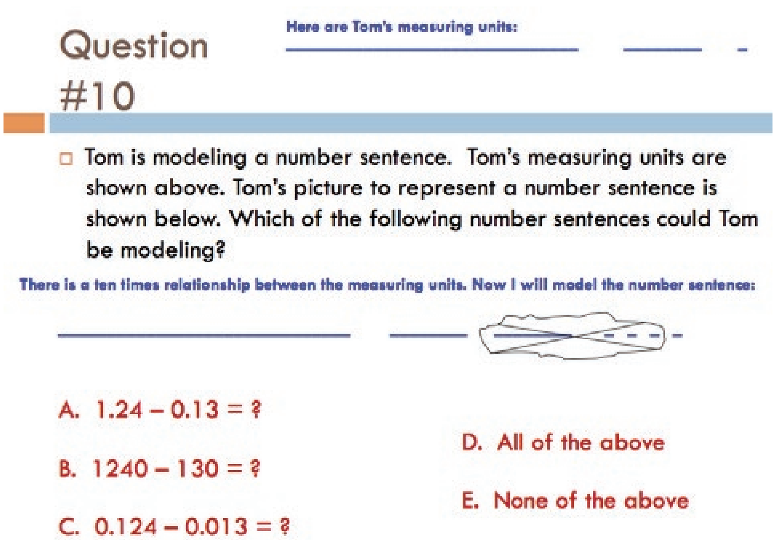
1. Names of the instructors: . . .

2. Semester and Year: . . .

3. Lesson Number, Lesson Name, and Activity Number where the change was made:

   Lesson 10, AddSubMeanII, Extra Homework

4. Please list the clicker item(s) that prompted the reported change to the lesson activity. Write the item(s) here, along with the multiple-choice answers.



5. What was the instructor group's rationale for making the change to the lesson activity? Please explain how the rationale is connected to the clicker items listed in #4.

   Pre-service teachers struggle with interpreting student work when diagrams are missing labels, measuring units, or are otherwise unclear. We thought that they needed to improve on their flexibility in how they interpret diagrams when information is missing. Moreover, we wanted PSTs to realize the importance of having all of this information on a diagram to make the diagram clearly linked to one number sentence.

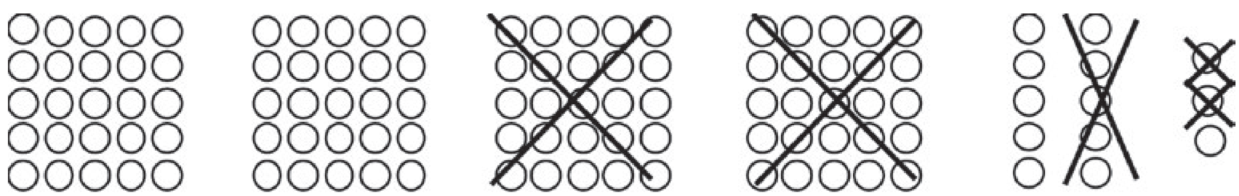
6. Please cut and paste the old version of the activity here:

   A previous version of this activity did not exist.

7. Please cut and paste the new version of the activity here:

   See attached document. [Here is a sample problem from the instructors' new instructional activities:]

   Consider the following diagram that Troy (a fourth grader) drew below. Does the diagram clearly illustrate a <u>*specific*</u> number sentence? If yes, state the number sentence. If not, explain why not.



8. Please provide the student results from each clicker item listed in #4 before and after you made the change(s) to the lesson activity.

Percent of students responding correctly:	Clicker Item 10
Before you made the change (i.e., last semester):	
Section 1	**38%**
Section 2	**23%**
Section 3	**5%**
Section 4	**71%**
After you made the change (i.e., this semester):	
Section 1	**82%**
Section 2	**76%**
Section 3	**87%**

9. Does the instructor group think the change was successful? Are the changes in clicker performance enough for the instructors to be satisfied or do they think there still is work to be done?

The instructors feel that the clicker item data provide evidence that there was a huge improvement in PSTs' ability to flexibly interpret diagrams for addition and subtraction of decimal numbers. Some improvement could be made to determine what specifically is the most difficult piece of interpreting diagrams with missing parts (i.e., is it the measuring units missing that makes it hard, is it the missing labels, or is it something else?). We think that we could be more proactive in linking a specific homework problem with a specific clicker item to measure more specifically what the PSTs have trouble with.

Instructors' Activity Change Report II

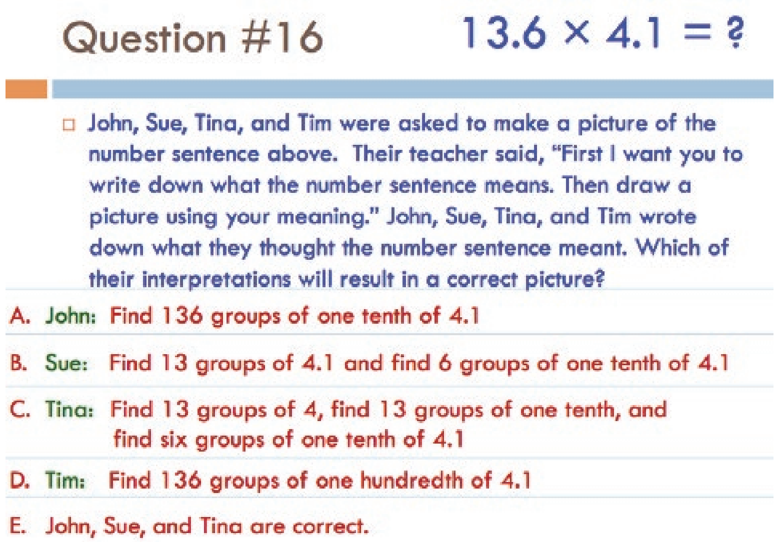
1. Names of the instructors: . . .

2. Semester and Year: . . .

3. Lesson Number, Lesson Name, and Activity Number where the change was made:

Lesson 14, MultMeaningI, Activity 2, Lesson 15, MultMeaningI, Extra Homework

4. Please list the clicker item(s) that prompted the reported change to the lesson activity. Write the item(s) here, along with the multiple-choice answers.



Question #19

- If you are solving 0.23 x 4.1 with a diagram on graph paper, you would get the right answer if you found the area of:

A. 23 groups of one-hundredth of 4.1

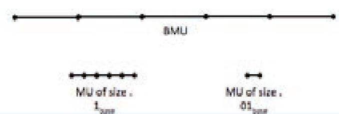
B. 2 groups of one-tenth of 4.1 and 3 groups of one-hundredth of 4.1

C. 2 groups of one-tenth of 4 and 2 groups of one-tenth of 0.1 and 3 groups of one-hundredth of 4 and 3 groups of one-hundredth of 0.1

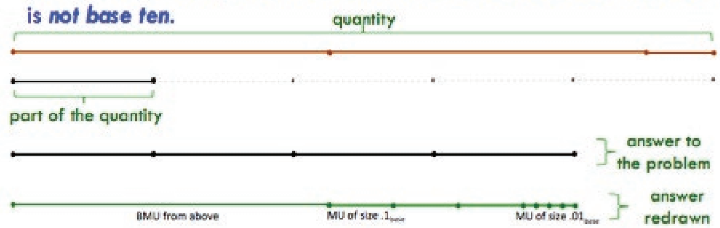
D. a, b, and c are all correct

E. Only a and b are correct

Question #20



- Using the measuring units as shown above, Michelle has drawn the picture below to represent and solve a number sentence in a base that is ***not base ten.***



What is the number sentence she is solving, and what is the answer?

A. $0.4_{fifteen}$ x $2.1_{fifteen}$ = ? and the answer is $1.34_{fifteen}$

B. 0.4_{five} x 2.1_{five} = ? and the answer is 1.34_{five}

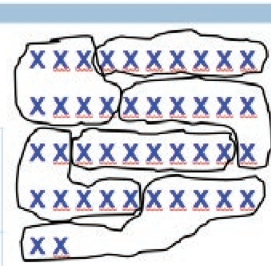
C. 4_{five} x 2.1_{five} = ? and the answer is 1.34_{five}

D. None above

E. Both b and c could be correct

Question #23

- Ted is modeling a division number sentence at the right. What number sentence could he be modeling?



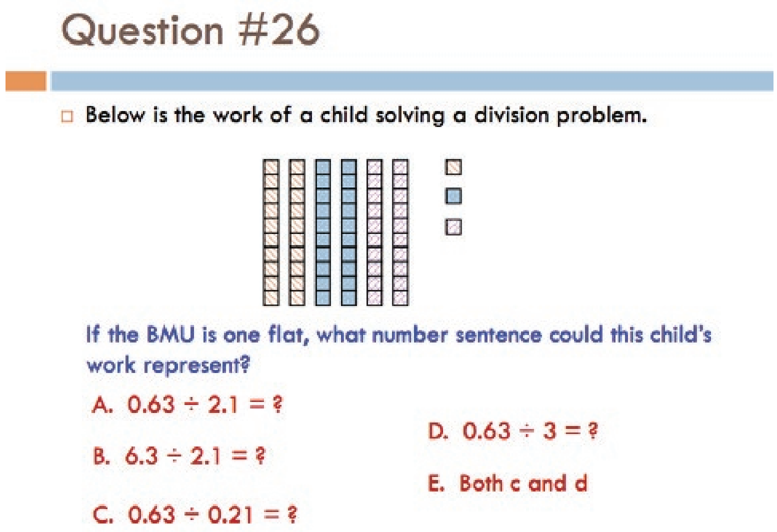
A. Ted could be modeling .42 ÷ .07 = ? using the repeated subtraction interpretation of division.

B. Ted could be modeling .42 ÷ 6 = ? using the partitioning interpretation of division.

C. Ted could be modeling 4.2 ÷ 6 = ? using the repeated subtraction interpretation of division.

D. None of the above

E. Both a and b



5. What was the instructor group's rationale for making the change to the lesson activity? Please explain how the rationale is connected to the clicker items listed in #4.

    Pre-service teachers struggle with interpreting student work when diagrams are missing labels, measuring units, or are otherwise unclear. We thought that they needed to improve on their flexibility in how they interpret diagrams when information is missing. Moreover, we wanted PSTs to realize the importance of having all of this information on a diagram to make the diagram clearly linked to one number sentence. We added a similar extra homework for addition and subtraction. Now, we added an extra homework for multiplication in the hopes that they would be able to transfer these ideas to division as well.

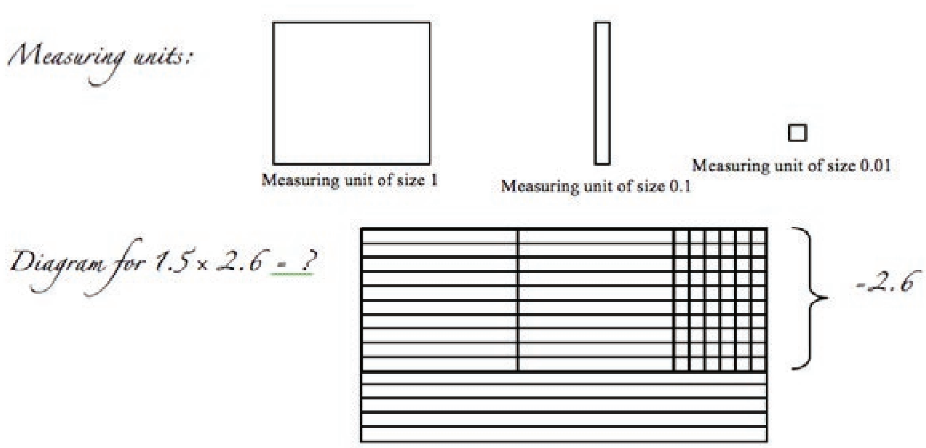
6. Please cut and paste the old version of the activity here:

    A previous version of these activities did not exist.

7. Please cut and paste the new version of the activity here:

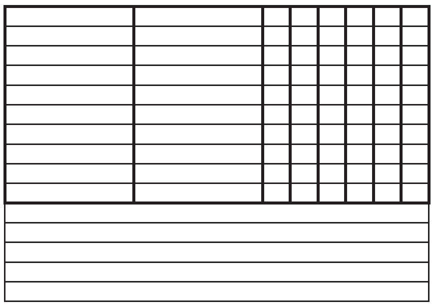
    For Activity 1, we added the following question to promote discussion of different ways to solve the multiplication problem: *What would be the all-at-once interpretation and the by-place interpretation if we broke up the 3.1 into place value parts?* For the extra homework, please see the attached document. [Here is a sample problem from the instructors' new instructional activities:]

Consider the following diagram for $1.5 \times 2.6 = ?$

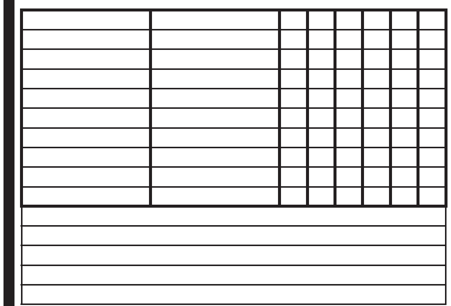


Three students say the following. Under each student's statement, show how to highlight what the student sees in the diagram.

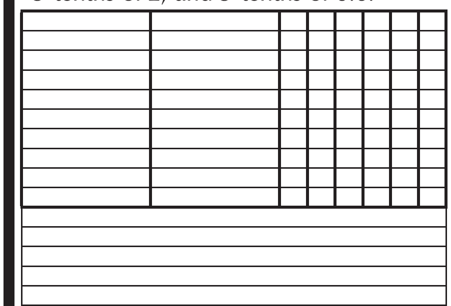
Pamela: I see 1 group of 2.6 and 5-tenths of 2.6!



Cory: I see 15-tenths of 2.6!



Sara: I see 1 group of 2, 1 group of 0.6, 5-tenths of 2, and 5-tenths of 0.6!



8. Please provide the student results from each clicker item listed in #4 before and after you made the change(s) to the lesson activity.

Percent of students responding correctly:	Clicker Item 16	Clicker Item 19	Clicker Item 20	Clicker Item 23	Clicker Item 26
Before you made the change (i.e., last semester):					
Section 1	**32%**	**76%**	**81%**	**14%**	**46%**
Section 2	**27%**	**66%**	**71%**	**33%**	**50%**
Section 3	**45%**	**52%**	**79%**	**53%**	**67%**
Section 4	**31%**	**81%**	**58%**	**24%**	**68%**
After you made the change (i.e., this semester):					
Section 1	**59%**	**100%**	**94%**	**61%**	**73%**
Section 2	**41%**	**77%**	**77%**	**43%**	**42%**
Section 3	**67%**	**73%**	**76%**	**61%**	**75%**

9. Does the instructor group think the change was successful? Are the changes in clicker performance enough for the instructors to be satisfied or do you think there still is work to be done?

The instructors feel that the clicker item data from clicker items 16 and 19 provide some evidence that there was satisfactory improvement in PSTs' ability to solve multiplication of decimal problems in more ways. We feel that the added discussion in Lesson 14 and some of the extra homework problems helped PSTs think about multiple ways to view multiplication. We still feel that improvements could be made to help PSTs understand the distributive property.

For clicker item 20, the instructors thought that the data showed slight improvement because of the extra practice that they had from the extra homework. But all of the extra practice was with an area model. We think that to improve what we did, discrete models and linear models should be included in addition to just the area model for multiplication.

For clicker items 23 and 26, the instructors thought that the data showed slight improvement, but we were not satisfied. We thought that the PSTs would be able to transfer the types of knowledge we discussed with interpreting diagrams for addition, subtraction, and multiplication to division diagrams. However, the data showed that this is not really the case. So, the instructors think it would be beneficial to have PSTs complete extra homework problems that focus on interpreting division diagrams with missing parts and where both partitioning and repeated subtraction are within the same diagram.

(**Return to page 135**)

Appendix B: Effectiveness Score Sheet

Please give a score to this lesson immediately after teaching it. Assign a score of 0 (low) to 2 (high) for each of the five criteria below. Indicate your score for each of the five criteria by checking the cell for 0, 1, or 2 in the table below. Then calculate the total score for the lesson by adding the five scores and write the score on the Total Score line below the table.

	Score		
	0 points (Not at all)	**1 point (Some, but needs improvement)**	**2 points (No improvement needed)**
The lesson succeeded in helping students achieve the learning goals.			
The lesson activities align well with the learning goals for the lesson.			
The lesson activities provided students with an opportunity to struggle with the critical mathematical concepts.			
The students had time to struggle with the critical mathematical concepts.			
During the lesson, the students received a clear explanation of the conceptual relationships among mathematical ideas, representations, and/or procedures from myself or other students.			

Total Score for Lesson x: ___________________

(Return to page 135)

A Comparison of Mathematics Classroom Observation Protocols

Melissa Boston
Duquesne University

Jonathan Bostic
Bowling Green State University

Kristin Lesseig
Washington State University—Vancouver

Milan Sherman
Drake University

In this article, we provide information to assist mathematics teacher educators in selecting classroom observation tools. We review three classroom observation tools: (1) the Reform-Oriented Teaching Observation Protocol (RTOP); (2) the Instructional Quality Assessment (IQA) in Mathematics; and (3) the Mathematical Quality of Instruction (MQI). We begin by describing each tool and providing examples of research studies or program evaluations using each tool. We then look across tools to identify each tool's specific focus, and we discuss how the features of each tool (and the protocol for its use) might serve as affordances or constraints in relation to the goals, purposes, and resources of a specific investigation. We close the article with suggestions for how each tool might be used by mathematics teacher educators to support teachers' learning and instructional change.

Key words: Classroom observation tools; Instructional quality; Reform-oriented teaching

In the September *Mathematics Teacher Educator* editorial, Smith (2014) describes how "tools" originally developed for research can be utilized by mathematics teacher educators to support teachers' learning and instructional change. Smith provides examples of how research frameworks for analyzing instructional tasks and teachers' questions can serve as scaffolds for teachers' learning (about cognitive demands or question types), instructional practice (in selecting tasks or asking questions), and reflection (on the nature of tasks or questions used during a lesson). In these examples, and more generally, tools highlight particular aspects of practice that research has identified as critical for enhancing students' learning of mathematics. Through the lens of a specific tool, teachers may be able to see aspects of instruction that previously blended into the myriad classroom activities occurring throughout a lesson. Once aspects of instruction are made visible, tools can provide a concrete structure for the development of new practices by specifying criteria and identifying standards for the implementation of the intended practice. Finally, tools can foster formative assessment and self-evaluation by focusing teachers' reflections on emerging or existing practices to identify strengths and/or motivate change.

In mathematics teacher educators' work as researchers, tools can serve to focus our analysis on key features of an intervention or treatment (e.g., professional learning activity, professional development initiative, or teacher education course or program). Tools can communicate standards for components of practice (e.g., mathematical tasks or teacher's questions) that can be shared across institutions, programs, research groups, and professional development settings. Tools allow us to gather data specifically related to the question under investigation (Smith, 2014), generating evidence that directly indicates the effectiveness or impact of the intervention and enables us to make valid claims about the intervention. In this way, tools can support evidenced-based practice in mathematics teacher education, particularly when the same tools are employed across multiple projects, sites, and investigators.

In this article, we take a closer look at a specific set of research tools that also hold promise for supporting teachers' learning and instructional change—classroom observation instruments. When the purpose of a pro-

This is an invited manuscript. In accordance with journal procedures, the manuscript was reviewed by the editor and two members of the Editorial Panel. It did not undergo a double-blind review process.

This manuscript is based on a presentation at the Association of Mathematics Teacher Educators (AMTE) 2014 Annual Meeting. The analysis of classroom observation tools described in this manuscript began within a working group in the Service, Teaching, and Research Program (STaR) in Mathematics Education in which the second, third, and fourth authors participated. Each of these authors contributed equally to this manuscript, and they are listed alphabetically.

fessional development project, intervention, or teacher education program is to change or impact some aspect of teachers' instructional practices, classroom observations are an essential component in identifying whether the intended practice is being implemented by teachers. The tools used to observe and assess classroom instruction must be able to identify the instructional practice(s) under investigation, using valid and reliable measures. Data and results generated by the classroom observation instrument should be useful in evaluating the effectiveness of the project, intervention, or program *and* for providing feedback to inform improvements in teachers' practice.

In this article, we provide information to assist mathematics teacher educators in selecting a classroom observation tool. We review three classroom observation tools and the protocols for their use:[1] (1) the Reform-Oriented Teaching Observation Protocol (RTOP); (2) the Instructional Quality Assessment (IQA) in Mathematics; and (3) the Mathematical Quality of Instruction (MQI). We have purposefully selected three tools with different foci and methods of use in order to raise issues regarding the selection of classroom observation tools more broadly. We begin by describing each tool and providing examples of research studies or program evaluations using each tool. We then look across tools to identify each tool's specific focus, and we discuss how the features of each tool (and the protocol for its use) might serve as affordances or constraints in relation to the goals, purposes, and resources of a specific investigation. We close the article with suggestions for how each tool might be used by mathematics teacher educators to support teachers' learning and instructional change.

The Reformed Teaching Observation Protocol

Purpose and Theoretical Foundation

The Reformed Teaching Observation Protocol (RTOP; Sawada et al., 2002), created originally to gather data for the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT), is a classroom observation protocol designed to measure the degree to which mathematics and science teaching are reform-oriented. Sawada and colleagues use the term *reform-oriented* to synthesize three instructional aspects: standards-based teaching,[2] an inquiry orientation in lesson design and implementation, and student-centered teaching practices. This protocol, grounded in a constructivist view of teaching (von Glasersfeld, 1989), builds upon reform efforts and advancements in K–12 standards for mathematics and science instruction (American Association for the Advancement of Science [AAAS], 1989; National Council of Teachers of Mathematics [NCTM], 2000; National Research Council [NRC], 1996).

The primary goal for the RTOP was to support reform efforts in professional development and teacher education. This goal demanded an instrument that would not only evaluate mathematics and science teaching but also help improve instruction (e.g., to provide a formative assessment of classroom instruction). In other words, evidence from an evaluation could communicate what went well during instruction, what could be improved, and how this improvement might occur.

The RTOP Instrument

The RTOP is a 25-item Likert-scale questionnaire (sample rubric provided in **Appendix A**; link to entire RTOP provided in **Appendix D**) examining three factors within the learning environment (subscales in parentheses): *Lesson Design and Implementation*, *Content* (propositional and procedural knowledge), and *Classroom Culture* (communicative interactions and student/teacher relationships). A brief description of the factors and subscales is offered here; Figure 1 provides example items to illustrate each subscale.

Lesson design and implementation, the only factor without subscales, identifies ways a teacher designs and sequences a lesson to support meaningful learning. The second factor, *Content*, includes two subscales: (1) *propositional knowledge* assesses whether instruction focuses on understanding core ideas meant to build conceptual understanding, and (2) *procedural knowledge* assesses how students solve problems and engage in problem-solving behaviors. Items on the propositional and procedural knowledge subscales are broad in nature and are not content specific. The third factor, *Classroom culture*, also has two subscales: (1) *communicative interactions* identify discourse moves that occurred during instruction, and (2) *student/teacher relationships* capture teacher-moves that facilitate a caring and nurturing environment.

The RTOP uses five-point Likert scales for each item. Raters are provided space for field notes, which supple-

1 We use the terms *classroom observation instrument* and *classroom observation tool* synonymously to refer to the set of rubrics used in classroom observations. We use the term *protocol* to encompass the rubrics and the requirements for their use (e.g., rater training, how they are used during an observation, etc.).

2 Given the cross-disciplinary nature of the RTOP, "standards-based teaching" is intended to highlight process standards for doing mathematics and engaging in science rather than content-specific standards.

RTOP factors and subscales	RTOP sample items
Lesson Design and Implementation	
(5 items)	"In this lesson, student exploration proceeded formal presentation."
Content	
Propositional knowledge (5 items)	"The lesson promoted strongly coherent conceptual understanding."
Procedural knowledge (5 items)	"Students made predictions, estimates, and/or hypotheses and devised means for testing them."
Classroom Culture	
Communicative interactions (5 items)	"The teacher's questions triggered divergent modes of thinking."
Student/Teacher relationships (5 items)	"The metaphor 'teacher as listener' was very characteristic of this classroom."

Figure 1. **Examples of RTOP items for each subscale.**

ment the quantitative scores and provide evidence to be used for shared reflection with the observed teachers. On the 0–4 scale, a score of 0 indicates that the item was "never observed" during the lesson, a score of 2 indicates that the item was observed at least twice, and a score of 4 indicates that the item was "very descriptive" of the lesson. A total of 10 or greater on any subscale suggests evidence of reform orientation for the construct assessed by that subscale. The RTOP's internal consistency is exceptionally high, $\alpha = 0.97$, indicating that individual teachers tend to score similarly between subscales. The subscales also have sufficient internal consistency (e.g., teachers tend to score similarly on the five items within a subscale), ranging from $\alpha = 0.80$ to as high as $\alpha = 0.93$, surpassing the threshold for reliable use in research settings (Gall, Gall, & Borg, 2007). Subscales may be used in place of the entire instrument if desired.

Administration of the RTOP

Researchers can use the RTOP to assess videotaped or live mathematics or science instruction in grades K–12, community college, or university settings. More observations of the same teacher increase the statistical validity of the instrument at the teacher level; however, one observation with two or more raters is sufficient to draw conclusions about an instructional episode. Prior to using the RTOP, raters are expected to complete a free online training (link provided in Appendix D), which takes approximately 1–2 days and should be completed by teams of at least two raters.[3] The online modules include (a) understanding the design of the RTOP and (b) reaching adequate interrater reliability with established raters. During the training, trainees evaluate recorded instruction and compare their scores with benchmark scores from RTOP developers. To be considered certified to use the RTOP in formal research, trained raters should have overall scores within +/– 5 points of the developers' scores, with each item score varying 1 point or less (Sawada et al., 2002). In research, two or more trained raters are expected to observe each lesson and then draw agreed-upon conclusions about the observed lesson. Note that the online training videos feature college-level education courses, and additional practice coding videos from K–12 mathematics classrooms may be necessary before using the RTOP.

Interpretation of RTOP Results

The RTOP generates an overall total score for each observed lesson, created by summing scores on all items within subscales. Scores range from 0 to 100, with higher scores indicating greater reform orientation. In a validation study including 141 Grade 6–16 mathematics and science teachers, the average RTOP score was 51.3 with a standard deviation of 20.1 (Sawada et al., 2002). An overall score of 50 (e.g., 25 items x 2 points per item) is the minimum threshold for considering instruction to have elements of reformed teaching (Sawada et al., 2002). For example, a score of 60 would suggest that the observed lesson was reform-oriented. Deeper inspection of the findings (e.g., by looking at scores for individual factors,

3 In-person training is no longer available.

subscales, or items) would be necessary to determine whether there was a general trend of reform-orientation across all factors or noticeable differences between factors (e.g., a total score of 10 or greater on any subscale suggests evidence of reform-oriented teaching for those specific constructs).

Quantitatively, RTOP scores can be used to assess change over time (e.g., by comparing teachers' pre/post scores), to assess the impact of an intervention or program (e.g., comparing control and treatment groups), or to provide a general measure of reform-oriented teaching in a school or district (e.g., percent of teachers above the 50-point threshold). Qualitatively, RTOP results are intended to foster discussions with individuals or groups of teachers about what raters noted during observed lessons and ways that mathematics or science teaching might be revised to promote reform-oriented teaching practices. In this way, the RTOP can be used in a school or district as a vehicle to spur building-wide or district-wide reform initiatives or in university settings to promote mathematics and science preservice teachers' ideas and practices regarding reform-oriented teaching.

The RTOP as a Research Tool

The RTOP has been used in several studies of mathematics and science teaching (Adamson et al., 2003; Dunleavy, Dede, & Mitchell, 2009; Jong, Pedulla, Reagan, Salomon-Fernandez, & Cochran-Smith, 2010; Roehrig & Kruse, 2005), and it has been adapted to meet specific needs in other studies (e.g., Ciancolo, Flory, & Atwell, 2006; Morrell, Wainwright, & Flick, 2004; Wainwright, Morrell, Flick, & Schepige, 2004). We present two examples here of how the RTOP was used to identify reform-oriented instructional practices.

Examining secondary instruction. Adamson and his ACEPT team (2003), the developers of the RTOP, conducted research to identify the impact of university coursework led by ACEPT instructors on secondary mathematics and science teachers' use of reform-oriented instructional practices. Two research questions guided this study: (1) Was there a difference in RTOP scores between teachers previously enrolled in ACEPT courses and those who did not experience such ACEPT courses? and (2) In what ways did students' content knowledge differ when comparing Grade 6–12 students taught by graduates from the ACEPT program and their peers taught by a non-ACEPT graduate?

ANOVA analyses identified a significant difference in mean RTOP scores between ACEPT graduates and comparison teachers, $F(2,26) = 3.44$, $p < .05$. Science students taught by ACEPT graduates had greater content knowledge than their peers taught by non-ACEPT graduates, $F(2,13) = 6.23$, $p = .01$, as measured by content-specific instruments. Thus, university instruction promoting reform-oriented teaching was associated with better teacher and student outcomes compared to a similar peer group.

Examining elementary instruction. Jong and colleagues (Jong, Pedulla, Reagan, Salomon-Fernandez, & Cochran-Smith, 2010) used the RTOP to examine classroom practices of preservice elementary mathematics teachers and to determine whether student teachers were enacting reform-oriented instruction after experiencing university coursework aligned with the vision of teaching shared by NCTM (2000) and the RTOP authors. Furthermore, the districts where student teachers were placed utilized reform-oriented curricula. The study investigated these questions: (1) To what extent did a sample of elementary preservice teachers implement reform-oriented teaching? and (2) What is the relationship between these teachers' levels of reformed instruction and their students' mathematical understanding (as measured by district-level mathematics tests)?

Results indicated that on average, the elementary student teachers engaged in reform-oriented mathematics teaching. Mean scores for the five subscales ranged from 1.94 to 2.42, which is near or above the minimum threshold (≥ 2) for instruction showing characteristics of reformed teaching. The overall RTOP average was above 50, indicating that in general, student teachers' instructional practices had characteristics of reform-oriented instruction. Subsequent analysis showed that three of the five RTOP subscales (propositional knowledge, procedural knowledge, and student/teacher relationships) were moderately correlated with students' mathematics understanding as measured by the district-level tests (i.e., $r \geq 0.51$). Thus, the student teachers appeared to be enacting the reform-oriented instructional practices advocated by their teacher preparation program, and these practices were associated with increased student learning.

Summary of RTOP

The RTOP has been used in several studies of mathematics and science teaching for 15 years (see Appendix D) and continues to be cited and utilized as a basis for new protocols (e.g., Morrell, Wainwright, & Flick, 2004). Studies using the RTOP share a vision that reform-oriented instruction is likely to lead to productive student outcomes. The RTOP plays an important role by instantiating this shared vision, which then allows observers and teachers to have a common ground for instructional

conversations within and between the content areas of mathematics and science. The RTOP generates quantitative data and also lends itself to offering meaningful feedback that can be conveyed to a teacher as a way to revise his or her mathematics or science instruction to align with the vision of reform-oriented instruction.

The Instructional Quality Assessment Mathematics Toolkit

Purpose and Theoretical Foundation

The Instructional Quality Assessment (IQA) Mathematics Toolkit (Matsumura, Garnier, Slater, & Boston 2008; Boston & Wolf, 2006) is a classroom observation protocol designed to measure the quality of mathematics instruction at scale using a combination of lesson observations, assignment collections, and student work. The IQA is based on two main constructs: *Academic Rigor* and *Accountable Talk*. The primary theoretical framework for Academic Rigor is Stein and colleagues' Mathematical Tasks Framework (Stein, Grover, & Henningsen, 1996), which considers the *cognitive demand* (i.e., type and level of thinking) that a mathematical task can potentially elicit from students, and how cognitive demands change throughout a lesson. Accountable Talk (Resnick & Hall, 1998) consists of the mathematical quality of classroom discourse with respect to accountability to the learning community and to the discipline of mathematics. Thus, the IQA assesses the quality of instruction based on the mathematical work that students *do* and *discuss* in the classroom, based on the cognitive demands and accountable talk moves observed during the lesson.

The IQA Instrument

Figure 2 provides an overview of the IQA constructs, rubrics, and indicators[4] (sample rubric provided in **Appendix B**; link to entire instrument provided in Appendix D). The construct of Academic Rigor contains three rubrics. First, *Potential of the Task* identifies the highest level of thinking and explanation that the written task has the potential to elicit from students. Second, *Task Implementation* measures the highest level of thinking in which the majority of students actually engaged during the observed lesson. Third, *Rigor of the Discussion* assesses the level of students' mathematical thinking and reasoning evident during a whole-class discussion following students' work on the task.

The *Rigor of the Discussion* rubric is a holistic assessment of students' mathematical representations, explanations, and strategies provided during whole-class discussion, whereas the Accountable Talk rubrics measure specific elements of classroom discussion at a finer grain size. *Accountable Talk* contains five rubrics: Participation, Teacher's Linking, Students' Linking, Teacher's Press, and Students' Response. *Participation* captures the proportion of students participating in the discussion. *Teacher's Linking* and *Students' Linking* capture accountability to the learning community, measured by the degree to which the teacher or students make connections to and build upon others' contributions to the discussion. *Teacher's Press* assesses the extent to which the teacher requires students to explain and justify their thinking, while *Students' Response* assesses the extent to which students provide such explanations or justifications. These rubrics

IQA construct	IQA rubric	IQA indicator
Academic Rigor	Potential of the Task	Instructional Tasks
	Task Implementation	Task Implementation
	Rigor of the Discussion	Explanations of Mathematical Thinking and Reasoning
Accountable Talk	Participation	
	Teacher's Linking	
	Students' Linking	
	Teacher's Press	
	Students' Response	

Figure 2. **An overview of the IQA constructs and rubrics.**

4 Rubrics for *Rigor of Teacher's Questions* and *Mathematical Residue* were added to the IQA in the fall of 2014 and are not discussed herein.

capture accountability to knowledge and rigorous thinking in the discipline of mathematics.

Each rubric is scored from 0–4 (with 0 indicating the construct is absent). For the Academic Rigor rubrics, low levels of cognitive demand (e.g., memorization or recall of facts or formulas, or the use of previously learned mathematical procedures without connections to concepts or meaning [Stein, Grover, & Henningsen, 1996; Stein, Smith, Henningsen, & Silver, 2009]) correspond to scores of 1 or 2, respectively. High-level cognitive demands (e.g., developing mathematical meaning for or with given procedures and/or open-ended problem solving [Stein, Grover, & Henningsen, 1996; Stein, Smith, Henningsen, & Silver, 2009]) that do not explicitly prompt students to explain their reasoning are assigned a score of 3, while those that do contain such prompts score 4. For the Accountable Talk rubrics, infrequent or formulaic talk moves (or student responses) score 1 or 2, and the presence and consistency of high-level talk moves score 3 or 4, respectively. As few as two observations per teacher may be sufficient to provide a reliable indicator of instructional quality when teachers comply with the data collection requirements (i.e., students engage in mathematical work followed by a whole-class discussion). When teachers were observed on consecutive days by the same (trained) observer, the IQA was found to have an acceptable internal consistency reliability (ϕ = .86) with only two observations (Matsamura, Garnier, Slater, & Boston, 2008). The dependability coefficient increased to .90 with one additional observation and to .94 for five observations.

Administration of the IQA

The IQA rubrics can be used in observations of K–12 mathematics classrooms. Researchers, professional development providers, and/or teacher educators can select individual IQA constructs or rubrics as aligned with the goals of the project, intervention, or program. A two-day face-to-face training is required for researchers to use the IQA; online training is not currently available. IQA developers intentionally designed the IQA rubrics to be used reliably during live classroom observations, but the IQA can also be used with videotaped lessons.[5] During the observation, raters take detailed field notes that are used to complete rubrics immediately following the observation. Once pairs of raters have achieved at least 80% exact-point agreement in the field (or using videotaped lessons from a project's dataset), lessons may be observed/scored by an individual rater.

Interpretation of IQA Results

When using the IQA, raters give the observed lesson one score on each IQA rubric. Because the IQA rubrics yield ordinal data, results from across an entire program, school, or project are reported as number (and percentages) of lessons at each score level for each rubric. Comparisons (e.g., between teachers' pre- and post-workshop data, data from control vs. project teachers, or different rubrics) should be conducted using nonparametric tests or tests of frequency or proportion. Descriptive data for each rubric, such as means and medians, are often reported to support interpretations of the results. For example, given the 4-point scale across all rubrics, a mean (or median) score above 2.5 is interpreted as an indicator of higher/consistent use of cognitively challenging tasks and/or accountable talk moves. Further, it is possible to make comparisons and identify relationships across different rubrics (e.g., comparing *Potential of the Task* to *Task Implementation* to determine whether high-level task demands were maintained) because the rubrics' score levels are similarly structured. The IQA score levels are also very descriptive, indicating specific characteristics or frequencies of instructional practice necessary for each score level. The detailed descriptors for each score level and the consistency in score levels across rubrics facilitate qualitative interpretations of the IQA results.

The IQA as a Research Tool

The IQA has been used to assess mathematics teachers' instructional practices in large-scale studies at the school or district level (e.g., Jackson, Garrison, Wilson, Gibbons, & Shahan, 2013; Quint, Akey, Rappaport, & Willner, 2007; Wilhelm, 2014) and professional development research (e.g., Boston & Smith, 2009, 2011; Sztajn, Wilson, Edgington, & Confrey, 2011). In this section, we describe two studies that used the IQA to identify reform-oriented instructional practices in secondary and middle level mathematics classrooms following teachers' participation in professional development or curricular reform efforts.

Measuring the effect of secondary mathematics teacher professional development. Boston and Smith (2009) used the IQA Mathematics Toolkit to assess the effectiveness of professional development designed to help teachers select and implement cognitively demanding tasks. Participants were 18 secondary mathematics teachers in the "Enhancing Secondary Mathematics Teacher Preparation" (ESP) project. Ten secondary mathematics teachers

5 A unique feature of the IQA is that it can be used with collections of student work *in lieu of* classroom observations. As the purpose of this paper is to discuss a sample of *classroom observation* protocols, the rubrics for student work collections are not discussed. See Matsumura, Garnier, Slater, & Boston (2008) for technical quality and Boston (2012) for examples of use.

who did not participate in ESP served as a control group. Instructional tasks, student work, and lesson observations were collected in fall, winter, and spring from ESP teachers and spring only from control group teachers. Only the IQA *Task Potential* and *Implementation* rubrics were used in this study because Accountable Talk was not a central feature of the ESP professional development.

Mann-Whitney tests indicated that ESP teachers' mean *Potential of the Task* score increased significantly from 2.54 to 3.01 from fall to spring ($z = 2.34$, $p < .01$ [one-tailed]). Chi-squared tests also indicated that the number of high-level tasks ($\chi^2(2) = 16.18$; $p < .01$) and implementations ($\chi^2(2) = 16.11$; $p < .001$) in ESP teachers' data collections increased significantly over time. While no significant difference existed between the control group and ESP teachers' fall lesson observations, ESP teachers' scores for spring lesson observations were significantly higher than scores of the control group for *Potential of the Task* ($z = 2.15$, $p = .02$ [one-tailed]) and *Implementation* ($z = 1.87$; $p = .03$ [one-tailed]). In addition, the results were independent of the type of curriculum teachers used, providing further evidence of the effectiveness of the professional development in supporting teachers to select and implement cognitively challenging instructional tasks.

Examining the Effect of Standards-based Curriculum and Professional Development. Boston (2012) used the IQA Academic Rigor and Accountable Talk rubrics to examine the effect of a district-wide curriculum adoption and accompanying professional development with 13 middle school mathematics teachers. Although lesson observations and assignments with accompanying student work served as data for the study, we focus here on the analysis of the lesson observations.

Results showed that for lesson observations following teachers' participation in the professional development initiative, the majority of tasks that teachers selected (58%) and implemented (65%) were low level (1 or 2 for *Potential of the Task and Implementation*). Only 27% of the whole-class discussions exhibited evidence of high-level thinking and reasoning (3 or 4 for *Rigor of Discussion*), while 54% were considered low level, and 19% of observed lessons lacked any discussion. Results were comparable for each Accountable Talk rubric, with minimal class discussions scoring 3 or 4 for *Teacher's Linking* and *Students' Linking*, or *Teacher's Press* and *Students' Response*.

These results demonstrated that teachers did not seem to be utilizing the high-level tasks provided by their *Standards*-based curriculum on a consistent basis for classroom instruction. Additionally, teachers were not consistently enacting whole-class discussions, and when they did, the discussion contributed little to students' opportunities to learn. Boston (2012) explained how these, and other, more fine-grained results based on the IQA, can be used to provide specific feedback to district leaders regarding the efficacy of the new curriculum adoption and professional development initiative. Furthermore, the results suggest pathways for improvement in teachers' instructional practice that can be addressed by ongoing professional developed tailored to the needs of the district.

Summary of the IQA

The IQA is a holistic assessment of mathematics instruction with a specific focus on the opportunities for students to engage in cognitively challenging mathematical work and thinking and to explain and express their reasoning in whole class discussions. Hence, the IQA rubrics are "best-suited for assessing reform-oriented instructional practices, for use in implementation studies of curriculum or professional development or to identify changes in the nature of school- or district-wide instructional practice over time" (Boston, 2012, pp. 95–96). The small number of observations needed to obtain a stable indicator of classroom practice allows for it to be used at scale and in settings in which videotape is not practical or possible. By identifying and measuring aspects of mathematics instruction that correspond to student achievement (Boaler & Staples, 2008; Stein & Lane, 1996; Stigler & Hiebert, 2004), the IQA provides feedback that can assist mathematics teacher educators in designing professional development that addresses specific elements of instruction (e.g., tasks, implementation, and/or discussion). Thus, the IQA can serve as both a lesson observation protocol and a tool for professional development.

The Mathematical Quality of Instruction

Purpose and Theoretical Foundation

The Mathematical Quality of Instruction (MQI) is a multidimensional assessment of the rigor and richness of the mathematics present during classroom instruction. The MQI must be used on videotaped instructional episodes (rather than live classroom observations). The instrument, designed by the Learning Mathematics for Teaching Project (**http://www.sitemaker.umich.edu/lmt/home**), was originally developed alongside efforts to conceptualize and validate measures of mathematical knowledge for teaching (MKT) (Ball, Thames & Phelps, 2008). MKT refers to the mathematical knowledge that is specifically entailed in the work of teaching. In accord with perspec-

tives advanced in the MKT literature, the MQI is designed to attend to the mathematics-specific components of the lesson and does not preference or measure a particular pedagogical approach. In other words, the authors claim that it is possible and beneficial to attend to the mathematical quality of an instructional episode regardless of the instructional methods. This orientation is reflected in the instrument's attention to *what* rather than *how* mathematical work is evidenced during the lesson.

A second key construct informing the design of the MQI is the instructional triangle (Cohen, Raudenbush, & Ball, 2003) in which instruction is conceptualized as interactions among teachers, students, and content. Given this more encompassing view of instruction, the instrument measures mathematical quality based on what a teacher says and does, what the students say and do, and what the curricula afford.

The original coding scheme was developed and refined through a synthesis of literature on mathematics classroom instruction and analysis of over 250 recorded lessons from 2nd– to 6th-grade classrooms. The codes are intended to capture elements of lessons that compromise the mathematical integrity of a lesson (e.g., the presence of errors or imprecise language), as well as aspects of instruction that support student learning (e.g., the use of multiple representations and explanations that focus on why something works). The instrument was revised in February 2014 to refine codes based on feedback from validation studies and to more explicitly align with mathematical practices outlined in the *Common Core State Standards for Mathematics* (CCSSM) (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

The MQI Instrument

The goal of the MQI is to measure the nature of the mathematical content available to students during instruction. To this end, the current instrument is organized around five dimensions of instruction: *Classroom Work is Connected to Mathematics*, *Richness of the Mathematics*, *Working with Students and Mathematics, Errors and Imprecision*, and *Common Core Aligned Student Practices* (see Figure 3; sample rubrics provided in **Appendix C**). Researchers can decide to use some or all of the MQI dimensions.

The six subscales within the *Richness of Mathematics* dimension capture the extent to which teachers and/or students (1) explicitly link and connect representations of mathematical ideas or procedures; (2) provide mathematical explanations that focus on why rather than how; (3) attend to the meaning of number relationships and operations; (4) discuss multiple procedures or solution methods; (5) develop mathematical generalizations based on examining instances or examples; and (6) fluently use mathematical language.

The dimension of *Working with Students and Mathematics* captures whether teachers can understand and respond to mathematical ideas students present and appropriately remediate student errors. The *Errors and Imprecision* dimension is more strictly focused on the teacher's use of correct, clear, and precise mathematical language and notation. The final dimension, *Common Core Aligned Student Practices*, incorporates the elements from what was called "Student Participation in Meaning-Making and Reasoning" in previous versions, along with additional subscales to denote the extent to which students work on contextual tasks and communicate about the mathematics. As a whole, subscales in this dimension measure student engagement in sense-making as indicated by the quality of student explanations; evidence of students' questioning, conjecturing, and generalizing mathematical ideas; and the cognitive demand of the task as it is enacted.

Videotaped lessons are chunked into equal intervals of 5 to 7.5 minutes, with each segment coded along the five dimensions. The first dimension, *Classroom Work is Connected to Mathematics*, is simply coded yes or no depending on whether at least 50% of class time (at least 3.75 minutes in a 7.5-minute segment) is connected to mathematics, rather than management or other activities. For the remaining four dimensions, raters take notes on each video segment and use these notes to score a number of subscales and the dimension overall after viewing the entire video (see Figure 3). Subscales and overall dimensions are scored using 4-point rubrics ranging from not present (0) to high (3). To illustrate, the rubric for the subscale *Linking Between Representations* and the dimension *Overall Richness of the Mathematics* are included in Appendix C. Researchers can opt to follow the "MQI Lite" protocol, where they provide an overall score for each dimension but do not score individual subscales.

Administration of the MQI

The MQI is designed for coding video of K–9 classroom instruction in mathematics. Video-based training is available free and online (link provided in Appendix D). Training requires approximately 16 hours that can be parsed into 1- to 2-hour increments. The training modules consist of detailed descriptions of codes and scoring guidelines, along with videos exemplifying different score points and practice tests. After working through these modules, individuals can become MQI certified by achieving an established percentage of agreement with master coding

MQI dimension	MQI subscales
Classroom Work is Connected to Mathematics	**1. Overall Classroom Work is Connected to Mathematics**
Richness of the Mathematics	2. Linking Between Representations 3. Explanations 4. Mathematical Sense-Making 5. Multiple Procedures or Solution Methods 6. Patterns and Generalizations 7. Mathematical Language **8. Overall Richness of the Mathematics**
Working with Students and Mathematics	9. Remediation of Student Errors and Difficulties 10. Teacher Uses Student Mathematical Contributions **11. Overall Working with Students and Mathematics**
Errors and Imprecision	12. Mathematical Content Errors 13. Imprecision in Language or Notation 14. Lack of Clarity in Presentation of Mathematical Content **15. Overall Errors and Imprecision**
Common Core Aligned Student Practices	16. Students Provide Explanations 17. Student Mathematical Questioning and Reasoning (SMQR) 18. Students Communicate about the Mathematics of the Segment 19. Task Cognitive Demand 20. Students Work with Contextualized Problems **21. Overall Common Core Aligned Student Practices**

Figure 3. **Dimensions and subscales in the MQI.**

on a selection of videos. The entire MQI protocol is made available once an individual has completed training and is consider certified.

Interpretation of MQI Results

In order to provide generalizable, reliable indicators of mathematical quality at the teacher level, the developers recommend that at least 3 lessons be scored independently by two coders. As described by Hill, Charalambous, and Kraft (2012), internal consistency reliabilities for overall dimensions substantially increased when three observations (per teacher) were scored by two certified coders compared to just one certified coder. When four observations were scored by two certified coders, internal consistency reliabilities continued to increase, but only slightly; for example, *Richness of the Mathematics* increased from 0.77 to 0.80, and *Errors and Imprecision* increased from 0.71 to 0.75.

A composite lesson score can be obtained by averaging scores from the five dimensions, and the composite scores from a teacher's four lessons can then be aggregated into one overall teacher-level score. At the teacher-level, the choice to report composite lesson scores, the overall scores for each dimension, or individual subscale scores is dependent on the purpose and grain-size of the research questions. Generally, the overall scores for each of the five dimensions are reported. However, scores for each subscale within a dimension can be used to provide formative feedback directly to teachers or aggregated to inform professional development. Scores from multiple teachers over the course of several observations may also be aggregated to provide feedback at the district

level. Such data can indicate trends across teachers or schools for program evaluation purposes (e.g., to measure impact of professional development or changes in the curriculum).

The MQI as a Research Tool

Two of the original goals prompting the development of the MQI were (a) to provide more than a propositional link between teacher knowledge and classroom instruction and (b) to capture the mathematical aspects of instruction as distinct from pedagogical strategy. Next, we discuss examples of recent studies designed to meet these goals and explore other potential uses of the MQI.

Relating teacher knowledge and classroom instruction. Hill and her colleagues at the University of Michigan (Hill et al., 2008) employed a mixed-methods approach to investigate how teachers' mathematical knowledge, as measured on paper-pencil MKT assessments, interacted with the mathematical quality of instruction. To facilitate the correlational study, MQI subscales were compressed into six separate dimensions (using an earlier version of the MQI). Teacher-level scores for each dimension and an overall lesson score were calculated by averaging scores across three videotaped lessons per teacher. Significant correlations were found between MKT and teachers' scores for the dimensions of *Total Errors*, *Language Errors*, and *Responding to Students Appropriately* (Spearman's $\rho = -.83$ and $-.80$ and $.65$, respectively). In addition to establishing strong positive correlations between MKT and MQI, qualitative analysis of both convergent (i.e., high MKT and high MQI) and divergent teacher cases revealed additional factors (i.e., curricular materials or teacher beliefs) that may mediate this relationship.

Building on this study, the MQI was used in a small-scale study to investigate the interrelationships amongst the mathematical quality of classroom instruction, curriculum, and teacher knowledge (Charalambous & Hill, 2012). Researchers aggregated overall scores on each MQI dimension across six videotaped lessons to categorize mathematical quality of instruction for each case-study teacher. Both MKT and the nature of the curriculum (e.g., reform-oriented or traditional) were positively related to teachers' abilities to use representations, provide explanations, and use precise language and notation as identified by the MQI.

Relating classroom instruction and students' mathematical achievement. MQI is one of five classroom observation protocols used in the larger Measures of Effective Teaching (MET) project[6] (Kane & Staiger, 2012). Preliminary findings indicate that the instrument does indeed measure qualities of instruction distinct from those assessed in non-subject specific observation tools (i.e., Classroom Assessment Scoring System [CLASS] or Framework for Teaching [FFT]). Specific studies within this project (see Appendix E) also reveal aspects of the MQI that are strongly or weakly correlated to teachers' value-added scores. Used at scale in this manner, the MQI can provide a vision of current mathematics instruction at a district or school level or indicate areas of instruction having the most impact on student achievement. For example, a comparison of scores across the five MQI dimensions indicated that the majority of mathematics lessons observed as part of the MET project were on-topic and relatively error free. However, the lessons were not necessarily mathematically rich, and students were given few opportunities to engage in sense-making activity (Kane & Staiger, 2012).

Summary of the MQI

The MQI instrument provides a reliable, quantifiable measure of the mathematical quality of instruction and is useful in a variety of settings. Two goals prompting the development of the MQI were (a) to provide more than a propositional link between teacher knowledge and classroom instruction and (b) to capture the mathematical aspects of instruction as distinct from pedagogical strategy. It can support large-scale studies attempting to establish correlations between classroom instruction and factors such as teachers' mathematical knowledge, professional development experiences, curricular materials, or student achievement. The level of detail provided within each subscale can also support qualitative analysis necessary to identify aspects of instruction that are especially enhanced or constrained by such factors.

Looking Across Classroom Observation Instruments

The classroom observation instruments reviewed herein (RTOP, IQA, MQI) have all been validated for use in mathematics education research. Each instrument can also be used to support mathematics teacher education, professional development, and program evaluation. Furthermore, each tool has features that may strengthen or limit its use, depending upon the specific contexts, resources, and research questions under investigation in a given study. In this section, we highlight features of the three tools that may help mathematics teacher

6 The MQI Lite, used in this comparative study, provides only overall scores for the five dimensions of the MQI, rather than scores for each subscale.

educators select a classroom observation tool appropriate for analyzing the specific aspect of instructional practice under investigation. We do this by considering the focus and contexts, as well as the features of each tool that may serve as affordances or constraints in relation to the goals, purposes, and resources of a given study.

Focus and Contexts

We propose that the purpose of a classroom observation instrument is to support knowledgeable raters' ability to notice the same aspects of instruction—ideally, aspects of instruction that impact students' learning of mathematics. Each tool reviewed herein helps the observer notice specific aspects of an instructional episode. What is noticed prompts the observer to make inferences about, and create meaning for, the instructional episode as a whole and specific events within that episode. With the multitude of events occurring throughout a mathematics lesson, each tool focuses the observer's attention in ways that elevate the importance of some events and reduce the importance of others.

The RTOP focuses the observer's attention on general features of reform-oriented instruction. Though the RTOP was developed to apply to mathematics and science, the reform-oriented instructional practices assessed by the RTOP rubrics are not inherently content-specific and may be applicable to other content-areas beyond mathematics and science. The large number of specific prompts within subscales makes the RTOP ideal for providing feedback to teachers regarding reform-oriented instruction. As the RTOP was designed to identify reform-oriented instructional practices, researchers or teacher educators should have some indication or expectation of the presence of these practices in order for the RTOP to be an appropriate tool. Contexts in which the RTOP may be ideal include assessments of (a) school-wide reform initiatives across multiple content areas, (b) preservice teachers and programs at the elementary level or across multiple secondary education content areas, or (c) professional development focused on the specific constructs identified by the RTOP (see Figure 1). For example, the RTOP would be a good choice to provide data for the research questions "Are preservice teachers in our program able to enact reform-oriented instruction?" (Jong, Pedulla, Reagan, Salomon-Fernandez, & Cochran-Smith, 2010) or "Did secondary mathematics and science teachers utilize reform-oriented instruction after experiencing such practices during university coursework?" (Adamson et al., 2003).

The IQA draws observers' attention to specific aspects of reform-oriented mathematics instruction, namely, cognitively challenging instructional tasks, task implementation, and discussion (including accountable talk). As the IQA was designed to identify specific reform-oriented instructional practices, some indication or expectation that these practices exist, are valued, or are intended to be developed over time should exist in order for the IQA to be appropriate to use. Contexts in which the IQA may be ideal include school-wide mathematics reform initiatives, preservice mathematics teachers and programs (at the elementary or secondary education level), and professional development, curriculum implementations, or large-scale assessments of reform-oriented mathematics teaching, *specifically focused* on the constructs identified by the IQA (e.g., cognitive demand and discussion). While similar to the RTOP in its focus on reform practices, the IQA attends to mathematics instruction and the use of cognitively challenging instructional tasks. Studies ideally suited for the IQA might investigate questions such as "Does secondary mathematics teachers' implementation of cognitively challenging tasks improve following their participation in task-centered professional development?" (e.g., Boston & Smith, 2009) or "How are elementary teachers using cognitively challenging tasks provided in their curriculum?" (e.g., Quint, Akey, Rappaport, & Willner, 2007).

The MQI assesses the rigor and richness of the mathematics throughout a lesson. In contrast to the other two tools, the MQI does not privilege reform-oriented instructional practices (though such practices may generate higher scores on some MQI rubrics, such as *Working with Students and Mathematic*s, and *Common Core Aligned Student Practices)*. The MQI is thus appropriate to evaluate students' opportunities to learn mathematics across a variety of instructional approaches, regardless of whether there is an expectation for reform-oriented instructional practices. MQI developers specify its use for grades K–9, perhaps because of the focus on mathematical content and the demands this places on raters. Contexts in which the MQI would be ideal include professional development initiatives or curriculum implementation, preservice mathematics teacher education programs at the middle or elementary level, or large-scale assessments of mathematics teaching, with a focus on the quality of mathematics during the instructional episode. Questions well-suited for the MQI include "What is the relationship between teachers' mathematical knowledge and the mathematical quality of instruction?" (e.g., Hill et al., 2008) or "How does the mathematical quality of instruction relate to students' mathematical achievement?" (e.g., Kane & Staiger, 2012).

In summary, the tools presented here overlap in some aspects but selectively attend to general reform-oriented practices (RTOP), specific reform-oriented practices in mathematics instruction (IQA), or the mathematical quality

of instruction across a variety of instructional approaches (MQI). The choice of an observation tool should be driven by the alignment between the research questions in a given study and the aspects of instruction made salient by that particular tool. We discuss next how features of each instrument can serve as affordances or constraints, depending on the focus and resources of a given study.

Features as Affordances or Constraints

As presented in each review, each classroom observation instrument has its own protocol for use and interpretation. We note that any feature is not inherently good or bad but becomes an affordance or constraint *in relation* to the goals and resources of a particular study. While our discussion references three specific tools, the four affordances and constraints discussed herein can apply to classroom observation tools more generally.

First, consider the requirement for live and/or videotaped observations. Live or videotaped observations can be used for the IQA or RTOP, and videotaped observations are necessary for the MQI. Live observations can be less invasive, which may facilitate participation and consent from schools, teachers, students, and parents. Live observations eliminate the need for video equipment or technology to collect, store, and share videos; however, they may require greater human resources, as at least one trained rater is necessary to observe and code each lesson. In live observations, the observer has access to the teacher, students, instructional materials, and other artifacts of the classroom and school. The rater can gain a holistic sense of classroom events (e.g., listening to conversations in multiple small groups; capturing participation, which may be difficult to identify in a video), where results obtained from videotape are influenced by the camera view of the lesson. However, live observations produce only written records and artifacts of the lesson, whereas video provides the ability to accumulate records of practice, score lessons independently or as a group, and replay instances in order to better understand key components of the lesson. Through technology, video can be shared in ways that enable raters to be in different physical locations (e.g., collaborations between researchers at different institutions).

Second, consider the requirement for rater training. The MQI and RTOP provide free, online rater training, while IQA training is only available face-to-face from the rubric developer. MQI online training can be completed individually, and raters require a certain level of mathematics knowledge (and mathematical knowledge for teaching) to successfully complete the certification. This system is thorough and produces high-quality raters but perhaps limits the pool of potential raters who can achieve certification. RTOP training should be completed by the research group together to allow for discussion and consensus. Raters must have a general sense of reform-oriented practices; hence the potential pool of raters is larger than for the MQI. To use the IQA, research groups attend a training session provided by the developer. Raters develop the ability to classify tasks by level of cognitive demand (Stein & Smith, 1998), identify features of task implementation that serve to maintain or reduce cognitive demands (e.g., Henningsen & Stein, 1997), and identify specific features of discussions and accountable talk moves (e.g., Resnick & Hall &, 1998).

Third, consider how explicit each rubric is in providing descriptions of score levels. Rubrics can consist of detailed score levels (e.g., IQA and MQI) or more general, “sliding scale” score levels (e.g., the individual RTOP prompts), which create differences in using the instruments and interpreting the results. Using the RTOPs’ more general score levels, raters focus on the number of occurrences of each indicator. Results thus identify the presence or absence of reform-oriented practices and can serve to indicate which practices were lacking or not observed. The IQA (Appendix B) and the MQI (Appendix C) provide very descriptive score levels within each rubric or subscale. These score levels serve as explicit indicators of exactly what features of a construct were strong or need to be improved to achieve the next level (for raters and for providing feedback to motivate instructional improvements).

Fourth, consider the scale of the research project and the usability of the rubrics. For small-scale projects, the variety of RTOP prompts can provide rich descriptive data to foster conversations with teachers or evaluate teacher preparation programs or professional development workshops. The IQA and the MQI were designed to be used reliably by trained raters in large-scale studies. For the IQA, the limited number and narrow focus of the rubrics make the IQA useable regardless of the scale of the study. The systematic process for using the MQI (e.g., coding timed segments) contributes to its usability at scale. All of the instruments can provide detailed results and feedback at the subscale or rubric level. For any instrument, while summing or averaging across individual rubrics or subscales into a composite or overall score may be useful for research purposes (e.g., correlating observation results to student achievement data or tests of teacher content knowledge), collapsing subscales or constructs into a single score also reduces the level of detail and specificity for which the results can be reported and interpreted.

Figure 4 provides a summary of features of each instrument discussed herein. As presented, a given feature of a tool can serve as an affordance or constraint.

	RTOP	IQA	MQI
Focus	Reform-oriented instruction	Cognitively challenging tasks, implementation, and discussion	Mathematical quality of instruction
Contexts	Reform-oriented instruction: • Building-wide reform initiatives • Professional development initiatives • Preservice teacher education programs	Reform-oriented mathematics instruction: • Professional development initiatives • Curriculum implementation • Preservice mathematics teacher education programs • Large-scale assessments of reform-oriented mathematics teaching	Mathematics instruction: • Professional development initiatives • Curriculum implementation • Preservice mathematics teacher education programs • Large-scale assessments of mathematics teaching
Affordances	• General indicators of reform-oriented instruction • Can be used across content areas • Live or videotaped lessons • Teacher-level data • Validated in MET Study • Provides rich descriptive data for discussions with teachers • Can show change over time • Free on-line training	• Very specific focus; can make explicit connections to PD • Live or video-taped lessons, or student work • Can be used at scale; can provide school- or district-level data • Descriptive statistics reported on individual rubrics • Well-defined score levels, explicit about what is needed to achieve next score level • Inter-rater reliability • Validated in prior studies • Provides rich descriptive data for discussions with teachers • Can be used over time, across sites, or compared to prior research	• Not biased toward any type of instruction • Videotaped lesson observations • Can be used at scale; can provide school- or district-level data • Well-defined score levels, explicit about what is needed to achieve next score level • Inter-rater reliability • Validated in MET Study • Many rubrics: provides rich descriptive data for discussions with teachers • Can be used over time, across sites, or compared to prior research • Free online training
Constraints	• Not inherently mathematical • Many indicators—difficult to use at scale and to get exact-point agreement between raters • No descriptors for each score level; not explicit about what is needed to achieve next score level • Training videos do not depict K–12 instruction	• Limited focus • Bias toward reform-oriented mathematics teaching • Accessibility of training • Not appropriate for comparisons or evaluations when there is no expectation of reform-oriented mathematics instruction	• Limited to K–9 • Not intended for live observations • Raters need adequate MKT • Broader focus may make it harder to establish explicit connections to PD

Figure 4. **Summary of features of the RTOP, IQA, and MQI.**

Using the Classroom Observation Instruments to Support Instructional Change

The classroom observation instruments reviewed herein are often used to evaluate teachers' instructional practices and provide feedback to inform research studies or interventions, teacher preparation courses or programs, or professional development efforts. While designed as research tools, each classroom observation instrument can also provide important information to support teachers' learning and instructional change by (1) serving as tools used in professional development and (2) providing a focus for formative assessment or self-evaluation of practice. Each of these uses will be discussed briefly, with specific reference to the classroom observation tools.

First, the classroom observation instrument themselves can be used as tools in professional development or teacher education. Used in this way, the tools could support teachers to notice the aspects of instruction central to each rubric, provide criteria for analyzing the aspects of instruction, and communicate a standard or develop a shared vision for practice. The RTOP has been used as a professional development tool for instructional planning, teaching, and reflecting on instruction (e.g., Ciancolo, Flory, & Atwell, 2006; Lawson et al., 2002; MacIsaac & Falconer, 2002). Lawson and his team (2002) used the RTOP to develop and frame summer institutes aimed at supporting college-level instructors to design and implement reform-oriented instruction. The IQA rubrics assess instructional practices (e.g., tasks, implementation, and/or discussion) that can be fostered through professional development (Boston & Smith, 2009, 2011). In fact, a variety of professional development materials currently exist that engage teachers in learning about and analyzing tasks, task implementation, and discussion (e.g., Stein & Smith, 1998; Stein & Smith, 2011; Stein, Smith, Henningsen, & Silver, 2009). Professional development or teacher education activities could be developed to incorporate the IQA rubrics explicitly (similarly for the RTOP and the MQI), where teachers might use the rubrics to analyze tasks, curricula, or videos of classroom instruction. The IQA has been used in this way to design a professional development workshop for middle-school principals. The goal of the workshop was to enable principals to identify high-quality tasks, implementation, and discussion during informal observations in mathematics classroom and to provide formative feedback to mathematics teachers based on these observations (Boston, 2011; Boston, Henrick, & Gibbons, 2014). Similarly, because the MQI is designed for use with video, it could easily be incorporated into professional development opportunities or teacher education courses. In these settings, the MQI can raise teachers' awareness and understanding of critical components of quality mathematics instruction such as using precise mathematical language, linking representations, or focusing on patterns and generalizations.

Second, classroom observation tools can be used as tools for formative assessment or self-evaluations of practice. Teachers can use the tools to identify the presence and quality of specific practices and to provide concrete pathways for instructional change. Each tool has features that support formative assessment of practice. The RTOP identifies many specific items that allow for detailed feedback, and thus can be used to engage teachers in reflections or conversations about specific aspects of reform-oriented instruction. The IQA contains a small number of rubrics, but is designed with detailed score levels to indicate specific criteria for improving instruction within each rubric. For the MQI, the specificity provided in the rubrics, coupled with recent revisions to align the final dimension with CCSSM, can serve as a valuable learning tool as teachers begin to make sense of new standards. Clear examples and language can support teachers' reflection on how instruction (either their own or others') encompasses important mathematical practices. As a practice-based tool, the MQI can focus teachers' attention on key components of quality mathematics instruction such as using precise mathematical language, linking representations, or focusing on patterns and generalizations.

Summary and Conclusions

In this article, we have reviewed three classroom observation instruments to provide information about these tools and to present general considerations for selecting observation tools that would be useful to other mathematics teacher educators. The tools presented herein communicate and evaluate standards of practice in mathematics teaching and learning; specifically, of reform-oriented instruction (RTOP), the selection and implementation of cognitively challenging tasks and discussion (IQA), and/or the mathematical rigor and richness of a lesson (MQI). We propose that the selection of a classroom observation tool should depend upon the question under investigation and the focus (e.g., the aspects of instruction each tool helps you to notice) of a specific study, professional development project, or program evaluation. In this way, the tool generates evidence directly connected to the intervention, appropriate for assessing the impact or effectiveness of the intervention and establishing evidenced-based practice as stated in the call for *MTE* articles.

References

Adamson, S., Banks, D., Burtch, M., Cox, F., Judson, E., Turley, J., . . . Lawson, A. (2003). Reformed undergraduate instruction and its subsequent impact on secondary school teaching practice and student achievement. *Journal of Research in Science Teaching*, *40*, 939–957.

American Association for the Advancement of Science. (1989). *Project 2061: Science for All Americans: A Project 2061 report on literacy goals in science, mathematics, and technology*. Washington, D.C.: Author.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*, 389–407.

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. *Teachers College Record*, *110*, 608–645.

Boston, M. D. (2011). *The Instructional Quality Assessment as a learning tool for school leaders*. Presentation at the annual meeting of the Association of Mathematics Teacher Educators, Irvine, CA.

Boston, M. D. (2012). Assessing instructional quality in mathematics. *Elementary School Journal*, *113*, 76–104.

Boston, M. D., & Wilhelm, A. G. (in press). Middle school mathematics instruction in instructionally focused urban districts. Accepted by *Urban Education*, July, 2014.

Boston, M. D., Henrick, E. C., & Gibbbons L. (2014). *Enabling principals to support high quality mathematics instruction.* Presentation at the National Conference of Supervisors of Mathematics, New Orleans, LA.

Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education*, *40*, 119–156.

Boston, M. D., & Smith, M. S. (2011). A "task-centric approach" to professional development: Enhancing and sustaining mathematics teachers' ability to implement cognitively challenging mathematical tasks. *ZDM: International Journal of Mathematics Teacher Education*, *43*, 965–977. doi:10.1007/s11858-011-0353-2

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

Charalambous, C. Y., & Hill, H. C. (2012). Teacher knowledge, curriculum materials, and quality of instruction: Unpacking a complex relationship. *Journal of Curriculum Studies*, *44*(4), 443–466.

Charalambous, C. Y., Hill, H. C., & Mitchell, R. N. (2012). Two negatives don't always make a positive: Exploring how limitations in teacher knowledge and the curriculum contribute to instructional quality. *Journal of Curriculum Studies*, *44*, 489–513.

Cianciolo, J., Flory, L., & Atwell, J. (2006). Evaluating the use of inquiry-based activities: Do student and teacher behaviors really change? *Journal of College Science Teaching*, *36*, 50–55.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*, 119–142.

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology*, *18*, 7–22.

Henningsen, M. A., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, *28*, 524–549.

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, *41*, 56–64.

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

Jackson, K., Garrison, A., Wilson, J., Gibbons, L., & Shahan, E. (2013). Exploring relationships between setting up complex tasks and opportunities to learn in concluding whole-class discussions in middle-grades mathematics instruction. *Journal for Research in Mathematics Education*, *44*, 646–682.

Jong, C., Pedulla, J., Reagan, E., Salomon-Fernandez, Y., & Cochran-Smith, M. (2010). Exploring the link between reformed teaching practices and pupil learning in elementary school mathematics. *School Science and Mathematics Journal*, *110*, 309–326.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* (MET Project research paper, Bill and Melinda Gates Foundation). Available from http://www.metproject.org/downloads/MET_Gathering_Feedback_Practioner_Brief.pdf

Lawson, A., Benford, R., Bloom, I., Carlson, M., Falconer, K., Hestenes, D., Judson, E., . . . Wyckoff, S. (2002). Evaluating college science and mathematics instruction. *Journal of College Science Teaching*, *31*, 388–393.

Learning Mathematics for Teaching Project (2011). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education*, *14*, 25–47.

MacIsaac, D., & Falconer, K. (2002). Reforming physics instruction via the RTOP. *Physics Teacher*, *40*, 479–485.

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

Morrell, P., Wainwright, C., & Flick, L. (2004). Reform teaching strategies used by student teachers. *School Science and Mathematics Journal*, *104*, 199–212.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). *Common Core State Standards for Mathematics*. Washington, DC: Authors.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

Quint, J. C., Akey, T. M., Rappaport, S., & Willner, C. J. (2007). *Instructional leadership, teaching quality, and student achievement: Suggestive evidence from three urban school districts*. New York, NY: MDRC. Retrieved October 28, 2014, from http://www.mdrc.org/publications/470/execsum.html.

Resnick, L. B., & Hall, M. W. (1998). Learning organizations for sustainable education reform. *Daedalus*, *127*, 89–118.

Roehrig, G., & Kruse, R. (2005). The role of teachers' beliefs and knowledge in the adoption of a reform-based curriculum. *School Science and Mathematics*, *105*, 412–422.

Sawada, D., Piburn, Judson, E., M., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2000). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics Journal*, *102*, 245–253.

Smith, M. S. (2014). Editorial: Tools as catalysts for practitioners' thinking. *Mathematics Teacher Educator*, *3*, 3–7.

Smith, M. S., & Stein, M. K. (2011). *5 practices for orchestrating productive mathematics discussions*. Reston, VA: National Council of Teachers of Mathematics.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in reform mathematics project. *Educational Research and Evaluation*, *2*, 50–80.

Stein, M. K., & Matsumura, L. C. (2008). Measuring instruction for teacher learning. In D. Gitomer (Ed.), *Measurement issues and the assessment of teacher quality* (pp. 179–205). Thousand Oaks, CA: Sage.

Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection: From research to practice. *Mathematics Teaching in the Middle School*, *3*, 268–75.

Stein, M. K., Grover, B. W., & Henningsen, M. A. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*, 455–488.

Stein, M. K., Remillard, J., & Smith, M. S. (2007). How curriculum influences student learning. In F. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 319–369). Charlotte, NC: Information Age Publishing.

Stein, M. K., Smith, M. S., Henningsen, M., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction: A casebook for professional development.* New York, NY: Teachers College Press.

Stigler, J. W., & Hiebert, J. (2004). Improving mathematics teaching. *Educational Leadership*, *61*, 12–17.

Sztajn, P., Wilson, P. H., Edgington, C., & Confrey, J. (2011). *Learning trajectories and key instructional practices.* Paper presented at the 33rd Annual Conference of the North American Chapter of the International Group for the Psychology of Mathematics Education, Reno, NV.

von Glasersfeld, E. (1989). Cognition, construction of knowledge, and teaching. *Synthese*, *80*, 121–140.

Wainwright, C., Morrell, P., Flick, L., & Schepige, A. (2004). Observations of reform teaching in undergraduate level mathematics and science courses. *School Science and Mathematics Journal*, *104*, 322–335.

Wilhelm, A. G. (2014). Mathematics teachers' enactment of cognitively demanding tasks: Investigating links to teachers' knowledge and conceptions. *Journal for Research in Mathematics Education*, *45*, 636–674.

Authors

Melissa Boston, G14 Canevin, 600 Forbes Avenue, Duquesne University, Pittsburgh, PA 15282; bostonm@duq.edu

Jonathan Bostic, 529 Education Building, Bowling Green State University, Bowling Green, OH 43403-0001; bosticj@bgsu.edu

Kristin Lesseig, VUB 345, 14204 NE Salmon Creek Ave., Washington State University—Vancouver, Vancouver, WA 98686; kristin.lesseig@vancouver.wsu.edu

Milan Sherman, 211 Howard Hall, 2507 University Ave., Drake University, Des Moines, IA 50311;

milan.sherman@drake.edu

Appendix A: Sample Rubrics: Reformed Teaching Observation Protocol (RTOP)

IV.	CONTENT	
	Propositional knowledge	
6)	The lesson involved fundamental concepts of the subject.	0 1 2 3 4
7)	The lesson promoted strongly coherent conceptual understanding.	0 1 2 3 4
8)	The teacher had a solid grasp of the subject matter content inherent in the lesson.	0 1 2 3 4
9)	Elements of abstraction (i.e., symbolic representations, theory building) were encouraged when it was important to do so.	0 1 2 3 4
10)	Connections with other content disciplines and/or real world phenomena were explored and valued.	0 1 2 3 4
	Procedural Knowledge	
11)	Students used a variety of means (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent phenomena.	0 1 2 3 4
12)	Students made predictions, estimations and/or hypotheses and devised means for testing them.	0 1 2 3 4
13)	Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.	0 1 2 3 4
14)	Students were reflective about their learning.	0 1 2 3 4
15)	Intellectual rigor, constructive criticism, and the challenging of ideas were valued.	0 1 2 3 4

Note. A score of 0 indicates "never occurred." A score of 4 indicates "very descriptive."

(Return to page 155)

Appendix B: Sample Rubrics: Instructional Quality Assessment (IQA) Potential of the Task Rubric

<table>
<tr><td>4</td><td>The task has the potential to engage students in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships, such as:
Doing mathematics: using complex and non-algorithmic thinking (i.e., there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instructions, or a worked-out example); OR
Procedures with connections: applying a broad general procedure that remains closely connected to mathematical concepts.
The task must explicitly prompt for evidence of students’ reasoning and understanding.
For example, the task MAY require students to:
solve a genuine, challenging problem for which students’ reasoning is evident in their work on the task;
develop an explanation for why formulas or procedures work;
identify patterns and form and justify generalizations based on these patterns;
make conjectures and support conclusions with mathematical evidence;
make explicit connections between representations, strategies, or mathematical concepts and procedures.
follow a prescribed procedure in order to explain/illustrate a mathematical concept, process, or relationship.</td></tr>
<tr><td>3</td><td>The task has the potential to engage students in complex thinking or in creating meaning for mathematical concepts, procedures, and/or relationships. However, the task does not warrant a “4” because:
the task does not explicitly prompt for evidence of students’ reasoning and understanding.
students may be asked to engage in doing mathematics or procedures with connections, but the underlying mathematics in the task is not appropriate for the specific group of students (i.e., too easy or too hard to promote engagement with high-level cognitive demands);
students may need to identify patterns but are not pressed for generalizations or justification;
students may be asked to use multiple strategies or representations, but the task does not explicitly prompt students to develop connections between them;
students may be asked to make conjectures but are not asked to provide mathematical evidence or explanations to support conclusions</td></tr>
<tr><td>2</td><td>The potential of the task is limited to engaging students in using a procedure that is either specifically called for or its use is evident based on prior instruction, experience, or placement of the task. There is little ambiguity about what needs to be done and how to do it. The task does not require students to make connections to the concepts or meaning underlying the procedure being used. Focus of the task appears to be on producing correct answers rather than developing mathematical understanding (e.g., applying a specific problem-solving strategy, practicing a computational algorithm).
OR There is evidence that the mathematical content of the task is at least two grade-levels below the grade of the students in the class.</td></tr>
<tr><td>1</td><td>The potential of the task is limited to engaging students in memorizing or reproducing facts, rules, formulae, or definitions. The task does not require students to make connections to the concepts or meaning that underlie the facts, rules, formulae, or definitions being memorized or reproduced.</td></tr>
<tr><td>0</td><td>Students did not engage in a mathematical activity.</td></tr>
<tr><td>N/A</td><td></td></tr>
</table>

(Return to page 158)

Appendix C: Sample Rubrics: Mathematical Quality of Instruction (MQI) 4-point

Linking Between Representations

This code refers to teachers' and students' explicit linking and connections between different representations of a mathematical idea or procedure. To count, these links must occur across different representational "families" e.g., a linear graph and a table both capturing a linear relationship. So, two different representations that are both in the symbolic family (e.g., 1/4 and 0.25) are not candidates for being linked.

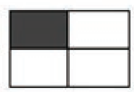
For Linking Between Representations to be scored above a Not Present:

- At least one representation must be visually present
- The explicit linking between the two representations must be communicated out loud

For Linking Between Representations to be scored Mid or High, two conditions must be satisfied:

- Both representations must be visually present
- The correspondence between the representations must be explicitly pointed out in a way that focuses on meaning (e.g., pointing to the numerator in 1/4, then commenting that you can see that one in the figure, pointing to the four in the denominator, pointing to the four partitions in the whole. "You can see the 1 in the 1/4 corresponds to the upper left-hand box, which is shaded, showing one piece out of four total pieces...")

$\frac{1}{4}$



For geometry, we do not count shapes as a representation that can be linked—we consider those to be the "thing itself." However, links can be scored in geometry if the manipulation of geometric objects is linked to a computation, e.g., showing that two 45-degree angles can be combined to get a 90 degree angle and linking that to the symbolic representation 45 + 45 =90.

Note: If links are made but underlying representation/idea is incorrect, do NOT count as linking between representations.

Not Present	Low	Mid	High
No linking occurs. Representations may be present, but no connections are actively made.	Links are present in a pro forma way; For example, the teacher may show the above figure and state that one quarter is one part out of four. These links will not be very explicit or detailed; both representations need not be present.	Links and connections have the features noted under High, but they occur as an isolated instance in the segment.	Links and connections are present with extended, careful work characterized by one of the following features: • Explicitness about how two or more representations are *related* (e.g., pointing to specific areas of correspondence) OR • Detail and elaboration about the relationship between two mathematical representations (e.g., noting meta-features; providing information about under what conditions the relationship occurs; discussing implications of relationship) These links will be a characterizing feature of the segment, in that they may in fact be the focus of instruction. They need not take up the majority or even a significant portion of the segment; however, they will offer significant insight into the mathematical material.

Overall Richness of the Mathematics			
This code captures the depth of the mathematics offered to students. Note: This is an overall code for each segment. It is not an average of the codes in this dimension, but an overall estimate of richness.			
Not Present	**Low**	**Mid**	**High**
Elements of richness are present but are all incorrect OR Elements of rich mathematics are not present.	Elements of rich mathematics are minimally present. Note that there may be isolated Mid scores in the codes of this dimension.	Elements of rich mathematics are more than minimally present but the overall richness of the segment does not rise to the level of a High. For example, a segment may be characterized by some Mid scores in the codes of this dimension or by an isolated High along with substantial procedural focus, etc.	Elements of rich mathematics are present, and either: a) There is a combination of elements that together saturate the segment with rich mathematics either through meaning or mathematical practices. OR b) There is truly outstanding performance in one or more of the elements.
Scoring Help - Overall Richness of the Mathematics			
In scoring Overall Richness, we assign a score of Not Present when there are no elements of richness present in the segment, or the components of richness that are present are all incorrect. For this code, we do not consider middling density of Mathematical Language to be an element of richness. That is, a segment could get a score of Low or Mid for Mathematical Language and still get a score of Not Present for Overall Richness.			

(Return to page 161)

Appendix D: Additional References and Resources

- This article is based on a session at the AMTE 2014 Annual Meeting. The AMTE session PowerPoint can be downloaded at: **http://padlet.com/wall/amte2014**
- Reformed Teaching Observation Protocol (RTOP)
- Training Site: **http://physicsed.buffalostate.edu/AZTEC/RTOP/RTOP_full/index.htm**
- Entire Protocol: **http://physicsed.buffalostate.edu/AZTEC/RTOP/RTOP_full/PDF/RTOPform_IN001.pdf**
- Research using the RTOP: Adamson et al., 2003; Dunleavy, Dede, & Mitchell, 2009; Jong, Pedulla, Reagan, Salomon-Fernandez, & Cochran-Smith, 2010; Roehrig & Kruse, 2005.
- Adaptations of RTOP: IOP in Ciancolo, Flory, & Atwell, 2006; see OTOP in Morrell, Wainwright, & Flick, 2004; Wainwright, Morrell, Flick, & Schepige, 2004
- IQA rubrics are available for viewing at: **http://peabody.vanderbilt.edu/docs/pdf/tl/IQA_RaterPacket_LessonObservations_Fall_12.pdf**
- For information on IQA training, contact Melissa Boston at: **bostonm@duq.edu**
- Research using the IQA: Boston, 2012; Boston & Smith 2009, 2011; Boston & Wilhelm, in press; Quint, Akey, Rappaport, & Willner, 2007; MIST Project: **http://peabody.vanderbilt.edu/departments/tl/teaching_and_learning_research/mist/index.php**
- Learning Mathematics for Teaching (LMT) Project: **http://www.sitemaker.umich.edu/lmt/home**
- Mathematical Quality of Instruction (MQI) Training Site: **http://isites.harvard.edu/icb/icb.do?keyword=mqi_training**
- Information on MQI studies based on the MET data is available at: **http://www.gse.harvard.edu/ncte/projects/core/default.php#.U6hZNqgozdA**

(**Return to page 155)**