2016 NCTM RESEARCH CONFERENCE
April 11–13 • San Francisco

Paper Title: Gaps between knowledge and practice: The case of teacher questioning
Author(s): Corey Webel
Session Title: Gaps between knowledge and practice: The case of teacher questioning
Session Type: Interactive Paper Session
Presentation Date: April 12, 2016
Presentation Location: San Francisco, California

Using simulated teaching experiences to develop preservice teachers' questioning practices

Corey Webel

Kimberly Conner

Introduction

The gap between knowledge and skill has long been recognized as a problem for teacher education (e.g., Borko et al., 1992), intensified in recent years by research revealing the complex and specific nature of knowledge for teaching (Hill, Ball, & Schilling, 2008). How can elementary teacher preparation programs, many offering only one or two courses for each content area, deliver the experiences needed to develop expertise in each content area? One approach is to help PSTs develop skills to *learn from* their teaching, with the expectation that much of what teachers need to know will have to learned on the job, but also that this learning will be more likely to occur if teachers develop certain reflective skills (Hiebert et al., 2007). Others suggest that teacher preparations focus on developing specific teaching skills by providing opportunities to practice the detailed work of teaching (Ball & Forzani, 2009).

Given the limited amount of time allotted for developing skill in mathematics teaching in teacher preparation programs, we propose that knowledge and skill might be developed simultaneously through simulated teaching experiences combined with focused reflection, organized into modules focused on specific teaching skills. We conjecture that technology might be used to standardize and study such experiences, reducing costs and providing wider access to learning opportunities.

In this paper, we report on an initial attempt to develop and test such a module. Using the online Lesson*Sketch* platform, we designed a set of simulated teaching experiences focused on the skill of posing questions in response to student thinking. We tested the module with 54 preservice teachers (PSTs) in the first of two mathematics methods courses in a university teacher preparation program. In this paper we report on the challenges we encountered, the successes that we observed, and our plans for improving the module in future iterations.

Theoretical Framework and Literature Review

Perspective on teaching

We see teaching as a complex set of practices that are undergirded by special kinds of knowledge that are unique to the work of teaching (Ball, Hill, & Bass, 2005; Hill et al., 2008). We also view teaching as highly situated, which means that "how a person learns a particular set of knowledge and skills, and the situation in which a person learns, become a fundamental part of what is learned" (Borko et al., 2000, p. 195). One implication of this view is that if preservice teachers are going to draw on the knowledge and skills that they gain in their education courses, the context of their learning experience needs to feel like teaching. Such experiences could include approximations that represent some authentic aspects of practice but also slow down the speed of classroom interactions and provide low-risk opportunities for novices to try, fail, and learn from their practice (Grossman, 2009).

Questioning practices

Questioning is clearly of interest to researchers who study the teacher's role in classroom discourse (Boaler & Staples, 2008; Hiebert & Wearne, 1993; Kazemi & Stipek, 2001; Stigler, Fernandez, & Yoshida, 1996). Research shows that the predominant focus of teacher questioning in mathematics classes has been to ascertain whether students can recall specific facts or carry out certain procedures (Hoetker & Ahlbrand, 1969). International comparisons in mathematics teaching have shown that these kinds of questions are especially prevalent in the United States (Givvin, Hiebert, Jacobs, Hollingsworth, & Gallimore, 2005; Kawanaka & Stigler, 1999; Stigler et al., 1996). A sequence of closed questions, intended to direct students through a series of procedural steps until they obtain the correct answer, has been referred to as *funneling* (Herbel-Eisenmann & Breyfogle, 2005; Wood, 1998). These types of questions position students as recipients of information rather than contributors to their own knowledge development (Boaler, 2003), and are unlikely to spur correct and complete explanations on the part of students (Franke et al., 2009).

Recent research has described questioning practices that, in contrast to funneling questions, are responsive to student thinking, drawing out and building on the specifics of students' ideas rather than imposing the teacher's idea (Jacobs & Empson, 2015; Kazemi & Stipek, 2001; Sherin, 2002). The National Council of Teachers of Mathematics' *Principles to Actions* (2014) advocates teacher questions that "build on, but do not take over or funnel, student thinking," and those that "make mathematical thinking visible" (p. 41). In this study, we are looking specifically at follow-up questions that teachers might pose immediately after eliciting an initial explanation about a student's solution. We define "high leverage questions" as having either of the following characteristics:

- Eliciting more information about a particular aspect of the student's expressed thinking (e.g., "can you tell me more about how you were thinking about the 1/3 here?")
- Drawing out student thinking in a way that has potential for the student to recognize a misconception, but doesn't directly imply that the student is wrong or impose the teacher's idea (e.g., "Can you show me the 5/6 of a sub that [your solution shows that] each person will get?")

Low leverage questions are those that have *any* of the following characteristics:

- Directly lead students to a preferred strategy or the correct answer ("why don't you try dividing each sub into three parts?")
- State or strongly imply that what the student has done is incorrect ("Is it 1/3 of a sub [as you've claimed], or 1/3 of half of a sub?")
- Presume that students have certain background knowledge, such as particular ways of talking about mathematical ideas (e.g, "What is the whole for your fraction?")
- Are general, not referencing specific aspects of the students' thinking (e.g., "Can you solve this another way?")

How do teachers learn to ask good questions? There have been efforts to explore methods for improving questioning practices. For example, Moyer & Milewicz (2002) found that the use of an analytic framework for reflecting on questions posed during one-on-one interview with students supported PSTs' in developing skill in questioning, though there was variation in the acquisition of this skill. For example, PSTs began to ask more follow-up questions, but in some cases only when students had incorrect answers. Nicol (1999) found that PSTs struggled to reconcile different purposes for questions, such as learning more about student thinking and also helping them arrive at a correct solution.

One hypothesis for improving teacher questioning is that if preservice teachers understand the mathematical concepts under consideration, they will be more likely to choose questions that support students in making sense of these concepts. There is research to suggest that, in general, teachers with higher levels of mathematical knowledge for teaching enact higher levels of practice (Campbell et al., 2014; Copur-Gencturk, 2015; Hill et al., 2008). However, the nature of the relationship between knowledge and practice is still underspecified.

Another hypothesis is that if teachers understand the general principles behind good questioning, such as those outlined in *Principles to Actions*, they will be able to choose questions that align with these general principles. However, when Spangler and Hallman-Thrasher (2014) attempted to help PSTs develop their ability to anticipate and respond to student thinking using imaginary task dialogues, they found that while PSTs were able to develop a repertoire of "standard" responses (e.g., "How did you get that?" "Can you tell me what you were thinking?"), they struggled to respond to students in ways that were specific to the problem at hand. This suggests that simply learning *about* good questions does not necessarily mean that teachers will be able to respond with such questions during mathematical conversations with students.

Finally, another hypothesis is that teachers simply need experience. The more they work with students, they better they will become at asking questions. However, the persistent lack of high quality questioning practices in the United States suggests that experience alone is not sufficient (e.g., Kawanaka & Stigler, 1999). One of the main purposes of this study is to explore what it might take to support preservice teachers in developing higher leverage questioning practices—in particular, whether and how teaching simulations might be useful.

Why a simulated experience?

A simulated experience is an *approximation of practice*, an opportunity "for novices to engage in practices that are more or less proximal to the practices of a profession" (Grossman et al., 2009, p. 2058). In this project, we used the Lesson*Sketch* platform to design online storyboard teaching scenarios, which include some aspects of the teaching context such as a classroom, students, a specific student and his work, dialogue (represented by text bubbles), etc. Lesson*Sketch* provides tools to aid in the reflection process; moments where the situation is paused and the user can be asked to make a choice, provide a comment, or ask a question. In contrast to interviews with real students (e.g., Moyer & Milewicz, 2002; Nicol, 1999) these features allow the designer a high degree of control over what the user can see, do, and notice within the representation (Herbst, Chazan, Chen, Chieu, & Weiss, 2011). Not only this, but because our simulation is online, it can be accessed easily by many participants and can generate a wealth of data in a short amount of time. Tweaks to the design can be made easily and new iterations can be subsequently tested with new populations.

Research question

Our primary goal in the project was to address the question, "What features of simulated experiences can support PSTs' development of skill in asking high leverage questions?" In analyzing the data, a number of interesting questions arose about how PSTs think about and enact questioning in response to student thinking. These included the relationships between their understanding of a mathematical task, their ability to interpret student work on that task, and their questioning practices in response to that student work. These relationships became secondary research objects that helped us understand how and why different features of the simulations were useful. In this paper, we describe our findings about these relationships, and also share the primary criteria PSTs invoked in their selection and evaluation of questions in the simulated teaching situations. We finish by describing some preliminary answers to main research question and our plans for refining the simulations to further our understanding of how simulations might support PSTs' development of questioning practices.

3 Methods

We used a design-based research approach (Brown, 1992), which focuses on the design of an instructional intervention and how alterations to the design over multiple cycles of implementation impact the learning experience of participants (Collins, Joseph, & Bielaczyc, 2004). We used this approach because we wanted to know not only whether students learned from the teaching simulations, but how and why they learned (or did not). In this paper, we share findings from the first iteration of our simulated teaching module.

3.1 Participants

This research project involved 54 preservice teachers, nearly all in their junior year of college, who were enrolled in a teacher preparation program at a large university in the Midwest region of the United States. Each PST was enrolled in one of five sections of an elementary mathematics content/methods course on numbers and operations, their first such course in a sequence of two. The course targeted fraction concepts for the first eight weeks of the semester, focused specifically on helping PSTs appreciate the role of the unit in constructing and naming fractions (Chval, Lannin, & Jones, 2013). Course assignments included explorations of mathematics with an emphasis on justification and reasoning, as well as analyzing and interpreting student work. As part of the program, each PST was assigned a field placement in an elementary classroom in which they spent at least 60 hours over the course of the semester.

3.1 Data Collection

The module described in this study was included in the required assignments for the course and consisted of three simulated teaching experiences set in the Lesson*Sketch* environment (in this paper referred to as the Emory, Matthew, and Brandon experiences), a reflection assignment, and a pre- and post-test (see Table 1). The pre-test was given before the start of the semester, and was presented to students through an online survey. It included a screencast component where PSTs used a tablet device to draw a solution and record a verbal explanation of their thinking. Then they analyzed a student solution to the same task. The post-test was identical to the pre-test, but included an additional example of student work.

Table 1: Data items collected

Item	Pre-post	Intervention (Experiences)	Focus of coding
1. Solution to task	X	X	Operating on quantities, justification, and correctness (score from 0-3)
2. Unpacking mathematical ideas in the task		X	Was an important subconcept identified?
3. Interpretation of student work	X	X	Was an important misconception identified?
4. Question composed (PST-constructed)	X	X	Does the question invalidate or funnel student thinking? Or does it draw out and/or build on student ideas?
5. Rationale for composed question	X	X	Open coding
6. Question(s) selected (from list)	X	X	[multiple choice, no coding]
7. Rationale for questions selected		X	Open coding
8. Evaluation of question after viewing student response		X	What criteria were invoked to support the evaluation?

Each teaching simulation experience included several probes to draw out the mathematical and pedagogical knowledge of PSTs. These included prompts for them to solve the mathematical task, describe the important mathematical ideas embedded in the task, and to interpret the thinking represented by the student work. In addition, PSTs were asked to compose a question that they would like to ask the student and provide rationale, and then to select a question from a list and provide a rationale for why they believed the selected question would be the best to ask the student. Finally, in each of the three Lesson*Sketch* experiences, the PSTs viewed a student response for the question that they selected, and were given an opportunity to evaluate their question once more, taking into account the student's response. In two experiences, PSTs were additionally provided the opportunity to "go back in time" to select a different question[1], to view the student's new response, and then determine which of the two questions they believed was "better" and explain why. In the third experience, PSTs were given one set of questions focused on "eliciting the student's thinking" (Brandon A) and then a set of questions focused on "moving Brandon's thinking forward" (Brandon B). Importantly, in the Brandon experience, PSTs did *not* directly compare the effect of the two questions.

We designed each experience to provide opportunities to select both high and low leverage questions. To help PSTs recognize the limitations of low leverage questions, we designed student responses to supply evidence that such questions do not help students move forward in their understanding (even if they lead to a correct answer). We also designed student responses that showed the affordances of high leverage questions, such as helping a student recognize a misconception or making visible more details about student thinking.

Analysis

Measuring knowledge. The three measures of knowledge (PSTs' solutions to the mathematical task, their unpacking of the task, and their interpretations of student work) were coded using the criteria shown in Table 1. For each category, codes were assigned by between two and four coders. We coded chunks of about 20% of the data at a time, checking percent agreement and discussing discrepancies as a group. When reliability reached 0.80, we divided the remaining data to be coded by individuals.

In the pre-post test, knowledge of the mathematical concepts was assessed through an analysis of the PSTs' screencast recordings of solutions to the Submarine Sandwich task (see Table 1). We created codes for whether PSTs operated primarily on symbols or quantities, what

[1] We designed the experience so that if PSTs originally selected a high leverage question, in the second round they would only see low leverage options, to ensure a contrast.

strategy they used to partition the sandwiches (if shown), what they identified as the unit for their solution, what justification they provided for their solution, and whether they obtained the correct answer. These codes were used to generate an overall score for each PST from 0 points (showing very little conceptual understanding) to 3 points (strong and consistent conceptual understanding). Interrater agreement was 86%,

In both the pre-post test and the experience, PSTs were asked, after viewing a piece of student work, to "explain how [the student] is thinking about fractions" and describe evidence from the student's work to back up their conclusions. For this prompt we coded whether the PST was able to identify a misconception that was central in the student's thinking, using a scale from 0 – 3 points (interrater agreement was 89%)

Measures of questioning practice. Questioning practices of the PSTs were measured on the pre-post test and in the Lesson*Sketch* experiences. In all of these, after viewing a piece of student work, PSTs were prompted to compose a question that they believed would be effective and provide a rationale for their question. We used our features of high leverage questions to code these.[2] In the pre-post test, specific features of questions (e.g., presence of funneling, specificity) were assigned a value and used to create a numeric score from 0-3 (interrater agreement was 89%).

After composing a question for the simulated student, PSTs then selected from a list of questions, and, in the experiences, viewed a student response to their selected question. Then they were asked to choose from the three options: 1) It was a good question; it accomplished what I wanted it to accomplish; 2) It was a good question, but [the student] didn't respond in the way I expected, or 3) It was maybe not the best question; I should have asked something

[2] In the pre-post test, specific features of questions (e.g., presence of funneling, specificity) were assigned a value and used to create a numeric score from 0-3 so that growth could be measured statistically.

different. Then they were asked to explain their choice. To analyze the evaluations, we developed a set of descriptive codes that will be described more in the results section. Interrater agreement on these codes was 80%.

4 Results

4.1 Relationship between content knowledge and PSTs' questioning practices

Our data showed little relationship between the understanding that PSTs demonstrated about a particular mathematical idea and their ability to either compose or select a high leverage question in response to a simulated student who demonstrated partial understanding of that idea. To the contrary, in some cases, PSTs with seemingly less understanding were more likely to pose a higher leverage question than those who appeared to have greater understanding.

4.1.1 Pre-post test. When comparing scores on the pre-and post-test, we saw improvements in content knowledge and interpreting student work, but no corresponding gain in either composing or choosing questions (see Table 2). We attribute the gains in the first two areas to PSTs' experiences in the methods course, which focused heavily on the mathematical ideas emphasized in the sub sandwich task (e.g., the importance of establishing the unit when naming a fraction). Questioning scores, however, did not show improvement, and in fact went down slightly.

Table 2: Gains from pretest to posttest

Component	Pre-test mean (SD)	Post-test mean (SD)	t-test significant?
Solution to task (scale from 0-3 points)	0.87 (0.89)	1.72 (1.00)	Yes, $p < 0.001$
Interpreting student work (scale from 0-3 points)	1.68 (0.91)	2.01 (0.94)	Yes, $p = 0.01$
Composing a question for Toby (scale from 0 – 3)	1.4 (0.55)	1.25 (0.64)	No, $p = 0.09$

Choosing questions (scale from 0-5)	2.65 (1.04)	2.58 (1.04)	No, $p = 0.46$

When comparing the *composed* questions on the pre and posttest, the total number of high leverage questions dropped from 18 to 13, and the number of low leverage questions rose from 32 to 41. When choosing questions from a list, there was some movement between the pre and the post, but little overall increase in the ratio of high to low leverage questions.

Content knowledge and questioning practices in the modules. In the Lesson*Sketch* experiences, we asked PSTs to identify the important mathematical ideas emphasized in the central task in the experience. In general, we saw little correlation between PSTs' ability to unpack the concept and their ability to compose a high leverage question. In the Emory experience, for example, 45 out of 54 PSTs identified an important mathematical concept embedded in the task. Of these, only 18 (40%) composed a high leverage question, compared to 3 out of the 8 (37.5%) who did not identify an important subconcept, not a significant difference.

Similar patterns emerged when PSTs were asked to select a good question from a list (see Figure 3). For example, in the Matthew experience, 30 PSTs were able to identify an important mathematical concept targeted in the task, but of these, 25 selected a low leverage question. Overall, 16 selected a high leverage question, and 11 of these were *not* able to identify an important mathematical concept. In this case, identifying a mathematical subconcept was *negatively* related to choosing a good question.

In the Brandon experience, only 20 PSTs were able to name an important subconcept of the task; many focused on describing the process of finding common denominators without any reference to fraction concepts that would support understanding of why the common denominator algorithm works. Of those 20, 13 (65%) chose a high leverage question on the first round, similar to the percentage of those who did not name a subconcept (58%). On the second round of questions, only 3 of the 20 PSTs who identified a subconcept selected a high leverage questions, compared to 1 of the 33 PSTs who did not identify a subconcept. In general, these results showed little relationship between the PSTs' ability to unpack the mathematical ideas in a task and their judgment about effective questions.

Interpreting student thinking and questioning practices in the simulations. In the teaching simulations, there were some situations in which there initially appeared to be a relationship between ability to interpret student thinking and questioning practice. For example, in Brandon A, 77% of the PSTs who gave a strong analysis of Brandon's work selected a high leverage question, most often "Can you tell me more about where the fourths, the sixths and the tenths are in your picture?" In contrast, of the 24 who focused on Brandon's need to understand how to carry out the common denominators procedure (considered an insufficient analysis), 58% selected the low leverage question, "Do you know what you need to do to the denominators before you can add fractions?" This initially suggested a relationship between these PSTs' ability to interpret student thinking and selecting high leverage questions.

However, this apparent relationship broke down on the next set of questions for Brandon. On that set, 38 PSTs chose the low leverage question, "In the problem it says that there are three fourths and three sixths. Are fourths the same as sixths?" including 17 of the 22 who sufficiently identified Brandon's misconception as being related to the size of the pieces. This suggests that PSTs were choosing questions *not* because of their potential to draw out or build on student thinking, but because of their potential to directly address what the PST assumed to be the student's misconception (in this case, not recognizing that fourths and sixths are different sizes and cannot be treated as tenths).

All of this data suggests that a robust understanding of the meaning of fractions and the ability to interpret student thinking does not, on its own, lead to improved questioning practices for PSTs.

What types of questions did PSTs tend to prefer and why?

To understand the lack of relationship described in the previous sections, as well as the general lack of apparent impact of our simulated student module, we looked more specifically at the criteria that PSTs invoked when evaluating their selected question *after* viewing the student's reaction. We found that a primary criteria for *evaluating the questions* (after seeing the students' response) was whether the question resolved the students' difficulty. PSTs tended to over-attribute understanding (or lack of understanding) to the student and credit or blame the question accordingly, or either ignore the student's response and simply discuss why they thought the question *should* have resolved the student's difficulty.

For example, in the Brandon experience, 70% of the PSTs chose the low leverage question "In the problem it says that there are three fourths and three sixths. Are fourths the same as sixths?" Brandon's response was, "Uh, no," with a confused expression. The teacher responded by saying, "Remember, you can't add pieces unless they are the same size." Brandon replied, "Okay." Of the 38 who selected this question and viewed the ensuring interaction, 34 believed it to be a good question. Of these, 14 seemed to read Brandon's response as implying understanding (e.g., "Now, Brandon remembers that you cannot add fractions unless they have equal parts. I feel like this question was very effective because of this point. He acknowledges that they are the same size parts so he shows some comprehension of the topic."). In 13 cases, PSTs acknowledged that Brandon was confused, yet continued to express appreciation for their original question on the basis that it *should* have resolved Brandon's confusion. For example, one PST explained, “Brandon seemed more confused than ever. I thought it was a very good question because it seemed like earlier everyone else in the class had a good understanding. I thought a simple reminder that the denominators had to be equal might be enough to jog his memory.”

In this experience we generally see PSTs evaluating their questions in terms of whether Brandon’s misconception was resolved. Some were prone to attribute understanding to Brandon’s response despite lack of evidence, while some continued to believe the question was good because it could or should have fixed his misconception even though they were not convinced he understood. Very few looked past the criteria of “fixing misconceptions” in their evaluations of their questions for Brandon.

From fixing misconceptions to understanding student thinking

In the Matthew experience, after selecting and evaluation the impact of their chosen question, PSTs were then asked to directly compare the results of a high leverage question and a low leverage question. After the first question, results looked similar to Brandon: Of the 37 PSTs who initially selected a low leverage question, 26 initially claimed that their question was good, 18 justifying their evaluation by claiming that the question targeted Matthew’s misconception. Only 11 of them claimed, after seeing Matthew’s response, that they should have asked something different, and 8 of these justified this by noting that the question did not fix Matthew’s misconception.

However, after seeing Matthew’s response to the second (high leverage) question, 10 of these 26 changed their minds, indicated that they preferred the high leverage question. Five of these explained that the high leverage question enabled them to get more insights into Matthew’s thinking. For example, one PST initially selected the question, “What if you think about both brownies as the whole? How much is each piece of the whole two brownies?" After seeing Matthew's response ("Each piece is a half"), the PST still judged the question to be good because "because it introduced the idea of the two brownies being one 'whole.'" After viewing the second question ("Susan said there was ¾ of the two brownies. Can you draw a picture of what you think that would look like?) and Matthew's response (drawing two brownies with ¾ of each brownie shaded), the PST decided that the high leverage question was better. She explained, "In the second question you got to see exactly what Matthew was thinking, which is different then what I had thought. Matthew's picture changed how I viewed his thinking/reasoning." This example shows a PST initially evaluating her question with respect to whether it targeted what she perceived to be Matthew's misconception. But after seeing the results of both questions, she attends to the fact that that the second question allowed her to see something new about Matthew's thinking. In other words, her criteria shifted from focusing on *fixing Matthew's misconception* to *understanding Matthew's thinking*.

This change seemed to be facilitated by the opportunity to see Matthew's response to both questions. Overall, after seeing both a high and low leverage question, 23 PSTs expressed preference for the high leverage question, while only 7 preferred the low leverage question (the rest believed the questions were the same). The number of PSTs who expressed appreciation for questions because they *targeted the students' misconceptions* dropped from 19 to 8 from Round 1 to Round 2, and the number of PSTs who expressed appreciation for questions *because they reveal student thinking* rose from 7 to 14. On the other hand, 14 PSTs disliked both questions, with 13 of these focused on the fact that Matthew did not change his mind (e.g., "Both of my questions didn't get the point I was trying to make across. Matthew still cannot see that two brownies can make up a whole"). These results show that while the design of the Matthew experience seemed to promote attention to more subtle aspects of student responses for many PSTs, others still seemed focused on the resolution of misconceptions.

Discussion and Conclusion

Our goal in creating the simulated teaching modules was to determine whether and how certain experiences might support PSTs in developing a particular teaching skill through experimenting with different decisions and reflecting on the effects of these decisions. One important finding was that, even though PSTs' content knowledge had improved and they were able to articulate general characteristics of "good" questions, they were still likely to select and defend low leverage questions in the simulated teaching situations. In this regard, our data revealed some of the challenges that teacher educators must overcome if graduates are expected to be able to enact complex teaching skills once they leave their teacher education programs.

The most significant challenge we encountered was PSTs' tendency to focus on "fixing" student's misconceptions. When we included cues to indicate that such questions did not resolve confusion (e.g., a confused expression, a hesitation represented by an ellipse, a clipped response in which the student simply agreed with the statement offered by the teacher, or in some cases, a direct admission of confusion on the part of the student), PSTs sometimes failed to notice them, attributing understanding when there was no evidence (Jansen and Spitzer, 2009). In other cases, they noticed the cues but took them to indicate that the question was not effective. Still another response to such a cue was to take note of it but to insist that the question *should have* resolved the students' confusion.

In the Matthew experience, not only did PSTs see the student's response to their question, they were required to compare the student's responses to both a high and low leverage question. In neither of the responses was the students' confusion resolved, but in the case of a high leverage question, Matthew revealed more about the nature of his misconception, providing potential leverage for a teacher to position him to recognize, make explicit, and confront his misconception. In this experience we saw direct shifts, as several PSTs moved from focusing on whether the question resolved Matthew's misconception to noticing how the high leverage question revealed useful information about his thinking. Still, some PSTs did not make this shift, maintaining their focus on whether confusion was resolved.

This research is limited in that we are only reporting on the initial iteration of the module. We do not yet know if revisions will result in improvements, although this baseline study increases our confidence that future gains will be attributable to the revisions to the simulations. Another limitation is that we are not assessing questioning practices outside of the simulation. Even if PSTs do show improvement within the simulated environment, we will not know whether these improvements will translate to improvements in practice, where decisions must be made in real time. However, the first step is showing improvement within the simulation. We believe that our first iteration has provided some valuable insights about the development of skill in teacher questioning, and we plan to continue to develop and study refinements to the simulations.

References

Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497-511.

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*(3), 14-22, 43-46.

Boaler, J. (2003). Studying and Capturing the Complexity of Practice--The Case of the "Dance of Agency". *International Group for the Psychology of Mathematics Education, 1*, 3-16.

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. *Teachers College Record, 110*(3), 608-645.

Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education, 23*(3), 194-222.

Borko, H., Peressini, D., Romagnano, L., Knuth, E., Willis-Yorker, C., Wooley, C., . . . Masarik, K. (2000). Teacher education does matter: A situative view of learning to teach secondary mathematics. *Educational Psychologist, 35*(3), 193-206.

Brown, A. L. (1992). Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*(2), 141-178.

Campbell, P. F., Nishio, M., Smith, T. M., Clark, L. M., Conant, D. L., Rust, A. H., . . . Choi, Y. (2014). The relationship between teachers' mathematical content and pedagogical knowledge, teachers' perceptions, and student achievement. *Journal for Research in Mathematics Education, 45*(4), 419-459.

Chval, K., Lannin, J., & Jones, D. (2013). *Putting essential understanding of fractions into practice in grades 3-5*. Reston, VA: National Council of Teachers of Mathematics.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design-based research: Clarifying the terms. Introduction to the learning sciences methodology strand. *The Journal of the Learning Sciences, 13*(1), 15-42.

Copur-Gencturk, Y. (2015). The effects of changes in mathematical knowledge on teaching: A longitudinal study of teachers' mathematical knowledge and instruction. *Journal for Research in Mathematics Education, 46*, 280-330.

Franke, M. L., Webb, N. M., Chan, A. G., Ing, M., Freund, D., & Battey, D. (2009). Teacher questioning to elicit students' mathematical thinking in elementary school classrooms. *Journal of Teacher Education, 60*(4), 380-392.

Givvin, K. B., Hiebert, J., Jacobs, J. K., Hollingsworth, H., & Gallimore, R. (2005). Are there national patterns of teaching? Evidence from the TIMSS 1999 video study. *Comparative Education Review, 49*(3), 311-343.

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.

Herbel-Eisenmann, B. A., & Breyfogle, M. L. (2005). Questioning our patterns of questioning. *Mathematics Teaching in the Middle School, 10*(9), 484-489.

Herbst, P., Chazan, D., Chen, C.-L., Chieu, V.-M., & Weiss, M. (2011). Using comics-based representations of teaching, and technology, to bring practice to teacher education courses. *ZDM Mathematics Education, 43*, 91-103.

Hiebert, J., Morris, A. K., Berk, D., & Jansen, A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education*, *58*(1), 47-61.

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

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372-400.

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 & Instruction, 26*(4), 430-511.

Hoetker, J., & Ahlbrand, W. P., Jr. (1969). The persistance of the recitation. *American Educational Research Journal, 6*(2), 145-167.

Jacobs, V. R., & Empson, S. B. (2015). Responding to children's mathematical thinking in the moment: an emerging framework of teaching moves. *ZDM Mathematics Education*.

Kawanaka, T., & Stigler, J. W. (1999). Teachers' use of questions in eighth-grade mathematics classrooms in Germany, Japan, and the United States. *Mathematical Thinking and Learning, 1*(4), 255-278.

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

Moyer, P. S., & Milewicz, E. (2002). Learning to question: Categories of questioning used by preservice teachers during diagnostic mathematics interviews. *Journal of Mathematics Teacher Education, 5*(4), 293-315.

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

Nicol, C. (1998). Learning to teach mathematics: Questioning, listening, and responding. *Educational Studies in Mathematics*, *37*(1), 45-66.

Sherin, M. G. (2002). A balancing act: Developing a discourse community in a mathematics classroom. *Journal of Mathematics Teacher Education, 5*, 205-233.

Spangler, D. A., & Hallman-Thrasher, A. (2014). Using task dialogues to enhance preservice teachers' abilities to orchestrate discourse. *Mathematics Teacher Educator, 3*(1), 58-75.

Stigler, J. W., Fernandez, C., & Yoshida, M. (1996). Traditions of school mathematics in Japanese and American elementary classrooms. In L. P. Steffe, P. Nesher, P. Cobb, G. A. Goldin, & B. Greer (Eds.), *Theories of Mathematical Learning* (pp. 149-177). Mahwah, NJ: Lawrence Erlbaum.

Wood, T. (1998). Alternate patterns of communication in mathematics classes: Funneling or focusing. In H. Steinbring, M. G. B. Bussi, & A. Sierpinska (Eds.), *Language and Learning in the Mathematics Classroom* (pp. 167-178). Reston, VA: NCTM.