The Teacher Responding Tool: Scaffolding the Teacher Practice of Responding to Student Ideas in Mathematics Classrooms

James P. Bywater

Dissertation Defense

Paper Three

Bywater, J. P., Chiu, J. L., Hong, J., & Sankaranarayanan, V. (2019). The Teacher Responding Tool: Scaffolding the teacher practice of responding to student ideas in mathematics classrooms. *Computers & Education, 139*, 16-30. doi: 10.1016/j.compedu.2019.05.004

Abstract

Research in teacher education highlights the importance of *responding to student ideas.* However, effectively noticing, interpreting, and then responding to students' mathematical ideas can be quite challenging for teachers as they try to balance multiple, competing goals in an authentic classroom setting. This study introduces the Teacher Responding Tool (TRT), and examines its role in scaffolding four high school teachers' responding practice. The TRT leverages natural language processing technology to provide teachers with automated, student-specific recommendations for how to respond to their student's ideas. By comparing teacher responding with and without the TRT recommendations, their interactions with the tool, their think-aloud data, and their post-project interviews, results demonstrate that the TRT recommendations helped teachers notice and respond to nuances in the mathematical ideas of their students. Implications for teaching and learning, responding tool development, and teacher professional development are discussed.

Keywords: adult learning, human-computer interface, improving classroom teaching, pedagogical issues, teaching/learning strategies

1. Introduction

Research in teacher education highlights the importance of supporting core teaching practices (e.g., Ball & Forzani, 2009; Grossman, Hammerness, & McDonald, 2009). In mathematics, *responding to student ideas* is considered a core, or high-leverage, teaching practice (McDonald, Kazemi, & Kavanagh, 2013; Lampert et al., 2013) because placing student ideas at the center of instructional decisions is critical for promoting equitable student participation, achievement, and agency (e.g., NCTM, 2014; NCTM 2018). *Responding to student ideas* is a complex practice, particularly when teachers need to make in-the-moment decisions. It involves teachers understanding both the normative and alternative mathematical ideas that a student may possess to infer what a student understands (Coffey, Hammer, Levin, & Grant, 2011). It also involves teachers prioritizing the mathematical ideas to which they respond, considering their importance for the student, for the whole class, and for reaching the intended learning goals (Ball, 1993). In addition, when responding to student ideas, teachers should ensure that their responses are student-specific, given in manageable chunks, do not make comparisons with other students, and do more than highlight errors (e.g., Kluger & DeNisi, 1996; Shute, 2008). Teachers should also be proficient with questioning strategies, or talk moves that support further student discourse (e.g., Chapin, O'Connor, & Anderson, 2009; Herbel-Eisenmann, Steele, & Cirillo, 2013).

Effectively noticing, interpreting, and responding to students' mathematical ideas can be quite challenging for teachers in the moment as they try to balance the multiple, competing goals in a classroom (Sherin, 2002). Given this complexity, teachers need opportunities to become skilled at responding to student ideas, and several approaches have been taken to help teachers learn these practices. Studies in teacher preparation have used a variety of "approximations of practice" (e.g., Grossman, Compton, et al., 2009) to help develop the skill of responding to student ideas. These approximations of practice include rehearsing a classroom interaction with peers (Lampert et al., 2013), enacting a short activity in an authentic classroom setting prior to a pre-service teacher placement (McDonald et al., 2013), and virtual classroom environments to enact and receive coaching on teaching practices (e.g., Ma et al., 2014). Studies focused on in-service teacher professional development have used teaching videos that can be paused, replayed, or tagged as the focus of teacher discussion groups (e.g., van Es & Sherin, 2002; Walkoe, 2015). Other studies have used wireless ear-pieces to hear live guidance from an expert outside the classroom who was watching and listening to the teacher as they interacted with students (e.g., Wake, Dailey, Cotabish, & Benson, 2017).

This paper examines an alternative approach to scaffolding the teaching practice of responding to student ideas that leverages natural language processing (NLP) technology. This approach builds upon research that effectively provides automated targeted written guidance directly to students (e.g., Aleven & Koedinger, 2002; Gerard, Matuk, McElhaney, & Linn, 2015) and instead provides guidance to teachers to help them write responses to students' explanations. The Teacher Responding Tool (TRT) automatically provides response recommendations for each student explanation. The recommendations are selected based on the mathematical ideas expressed by the student. The TRT does not automatically respond for teachers, but rather presents recommendations for teachers to consider, select, and edit. In the context of an authentic setting, we conjecture that providing automated recommendations to teachers with the TRT can help teachers develop their practice of responding to student ideas or more specifically, develop their practice of noticing student ideas, providing formative feedback to these ideas, and asking questions to develop student thinking. To examine this conjecture (e.g., Sandoval, 2014), we compare teachers responding in writing *with* and *without* the TRT recommendations, and ask research questions about the role of the TRT recommendations on the observable interactions of the teachers with the tool (research question 1), on the responding artifacts produced by teachers (research question 2), and on the teachers' perceptions about the development of their responding practice (research question 3):

1. How do teachers interact with the TRT recommendations?
2. To what degree do the teacher responses relate to their students' ideas?
3. How do the teachers describe the impact of the TRT recommendations on their responding practice?

2. Literature Review

Responding to student ideas involves teachers making instructional decisions based upon the ideas that students express. An initial component of responding is *noticing*, or identifying and interpreting the important ideas that students have (Jacobs, Lamb, & Philipp, 2010) so that teachers are able to decide how to respond based on these ideas. Responding to student ideas means that a teacher responds to the details of the ideas expressed by a student rather than responding by making personal comparisons with other students, or by making general statements that are not aligned with a student's ideas (Robertson, Atkins, Levin, & Richards, 2016). Research on *formative feedback* also provides insight as to how teachers can respond to student ideas (e.g., Hattie & Timperley, 2007; Shute, 2008). We examine the literature that relates to teacher discourse *"moves"* and *teacher questioning* because responding in ways that encourage students to express more of their ideas is necessary for future responding to occur (e.g., Chapin et al., 2009; McElhone, 2013; Herbel-Eisenmann et al., 2013). In addition, since this paper describes an intervention to *scaffold the practice of responding with technology*, we review the literature that relates to techniques and technologies that have been used to provide scaffolded opportunities for teachers to develop their practice of responding to student ideas (e.g., Dieker, Hughes, Hynes, & Straub, 2017; Wake et al., 2017; Walkoe & Levin, 2018).

2.1. Noticing

Noticing the ideas that students express is an initial component of responding to student ideas. Teacher noticing has been described broadly as what teachers identify as important during classroom learning; what connections teachers make between specific classroom events and broader pedagogical ideas; and what contextual information can be used to reason about the specific classroom events (van Es & Sherin, 2002). More specifically, the teacher noticing of student mathematical ideas involves a teacher identifying the important mathematical details of a student's strategy or explanation, attending to the subtlety and diversity of the ideas, making mathematical connections between them, and interpreting them in the context of the intended mathematics learning goals (Jacobs et al., 2010; Robertson et al., 2016). However, research demonstrates that noticing is difficult, and many teachers struggle to notice the diversity and nuance of student ideas in a classroom (e.g., Sherin, Jacobs, & Philipp, 2011).

Central to developing a teacher's expertise in noticing mathematical ideas is developing their Mathematics Knowledge for Teaching (MKT; Ball, Hill, & Bass, 2005). In addition to knowing commonly accepted solutions, MKT involves understanding the variety of non-normative approaches students are likely to have and knowing how to scaffold students from these ideas to the lesson goals (Ball, 1993; Ball, Sleep, Boerst, & Bass, 2009). One approach to build MKT with teachers has been Cognitively Guided Instruction (CGI; Carpenter, Fennema, Franke, Levi & Empson, 1999). CGI has been used to help teachers to map out the conceptual development of their student's mathematical thinking so that teachers are better able to make informed instructional decisions (Carpenter, Franke, & Levi, 2003). Teachers need opportunities to continue to develop MKT throughout their career.

A prerequisite for noticing student ideas is having mathematical tasks and activities that make students' thinking visible to teachers. For decades, mathematics educators have stressed the importance of students verbalizing, explaining, and creating mathematical arguments in classrooms (Countryman, 1992; Burns, 1995; NCTM, 2014). Despite evidence that student self-explanations of mathematical understanding can help students learn (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, & Wylie, 2014), the implementation of students verbalizing their ideas in mathematics classrooms is still limited (Teuscher, Kulinna, & Crooker, 2016). Technology-enhanced learning environments provide new opportunities for students to generate explanations of their understanding in mathematics classrooms (e.g., Web-based Inquiry Science Environment, Slotta & Linn, 2009; Stoyle & Morris, 2017). Such environments also have the potential to support teachers in responding to these student explanations.

2.2. Formative feedback

Responding to student ideas involves deciding what to say or do next based on an assessment of student ideas. Just because teachers notice student ideas does not necessarily mean that teachers can respond effectively to those ideas (Jacobs et al., 2010). This aspect of responding has much in common with the idea of formative feedback, which has been defined as information communicated to a student with the intention of modifying their thinking and improving learning (Shute, 2008). Giving feedback to students can have a positive impact on their learning, but this impact is dependent on how the feedback is given (Black & Wiliam, 1998). Feedback should address the features of a student response that relate to the lesson task, it should provide guidance that helps direct students towards the learning goals, and it should be provided in small enough pieces so that a student is more likely to consider the feedback (Shute, 2008; Narciss & Huth, 2004; Bransford, Brown, & Cocking, 2000). Feedback should avoid normative comparisons with other students, as well as references to the student as a person rather than their learning (Kluger & DeNisi, 1996). Feedback should not be accompanied with grades or excessive praise, and it should avoid interrupting students who are actively engaged (Shute, 2008; Wiliam, 2007). Feedback is effective at supporting student learning if it is able to guide a student from their current understanding to the desired learning goals (Hattie & Timperley, 2007).

2.3. Discourse "moves" and teacher questioning

Responding to student ideas also involves encouraging students to express more of their ideas so that future responding can occur. Teacher discourse "moves" and questioning strategies can also guide a student because they can encourage a student to think about ideas that will help them reach the desired learning goal. Models of teacher discourse moves such as *Talk Moves* (Chapin et al., 2009), or *Teacher Discourse Moves* (Herbel-Eisenmann et al., 2013), have identified responding practices that expert teachers make to further elicit a student's ideas. These moves are not limited to questioning strategies. For example, one move—revoicing—involves a teacher restating a student contribution but also asking the student to confirm that the teacher has correctly understood their idea. This interaction has the effect of highlighting the student idea but also ensuring that the student keeps the credit for the idea and that the student is positioned as someone with ideas worthy of discussion (O'Connor & Michaels, 1993). At the same time, many of these discourse moves such as pressing students to provide their reasoning, asking students to say more, or probing a student's thinking, involve teachers asking students questions in order to elicit the student's ideas (McElhone, 2013; Herbel-Eisenmann et al., 2013; Chapin et al., 2009).

The questions that teachers ask play a critical role in eliciting students' ideas, whether they are the ideas students initially have about a problem, or their ideas about another student's thinking (Franke et al., 2009). Given this importance, various classifications have been developed to highlight the differences between questions teachers ask and their role in eliciting student ideas. For example, Wood (1998) contrasts funneling and focusing questioning patterns, and Kazemi and Stipek (2009) compare low- and high-press questions. High-press questions are those that ask students to provide mathematical arguments rather than procedural descriptions, and to explore alternative ideas and strategies. Examples of high-press questions include: What evidence supports that? Can you say that in your own words? How did you figure that out? What will you do next? and Why do you think so? (McElhone, 2013). Questions have also been categorized as either assessing questions or advancing questions (Smith, Steele, & Raith, 2017). Assessing questions focus on what a student currently understands. They are intended to reveal the mathematics that a student understands, to elicit a student's problem-solving strategy, or to probe a student's thinking. Examples include: Why do you think that? What do others think about what ___ said? and How did you reach that conclusion? Advancing questions, on the other hand, are intended to build upon what the student currently understands and advance the student toward the learning goals. Examples of these questions include: Does that always work? Can you think of a counterexample? Do you see a pattern? and What would happen if…? Advancing questions ask students to reflect on their current understanding, and to extend or apply what they know to new situations (Smith et al., 2017).

2.4. Scaffolding the practice of responding with technology

In order to help teachers to develop expertise in responding to students' ideas, research suggests that teachers can build these skills with "decompositions" and "approximations" of this practice (Grossman, Compton, et al., 2009). Various forms of technology can help teachers engage with these approximations of practice. For example, reflection supported by teaching videos can support teachers in noticing and interpreting student ideas, and it affords teachers opportunities to discuss their interpretations of classroom interactions and develop their noticing skills together (van Es & Sherin, 2002; Walkoe, 2015). By participating in video-club professional development meetings, teachers have increased their attention to students' mathematical thinking, and the space they make in their classroom for this thinking to emerge (van Es & Sherin, 2010). Video and multimedia storyboarding tools such as Lesson*Sketch* have also been used to provide opportunities for teachers to develop their responding practice by focusing attention on the questions they might ask in particular situations (e.g., Walkoe & Levin, 2018). In their storyboard responses, teachers were found to improve their noticing skills but also ask better questions. This is important because Jacobs and colleagues (2010) found that becoming skilled at noticing does not necessarily mean that teachers use what they have noticed when responding. At the same time, more research is needed to find whether noticing and questioning skills developed in out-of-the-classroom settings transfer to face-to-face settings (Weston, Kosko, Amador, & Estapa, 2018).

One approach for supporting the transfer of responding skills developed in out-of-the-classroom settings is to enact these practices with face-to-face rehearsal (Lampert et al., 2013). Interactive, virtual, multi-user, avatar-based environments, such as *Second Life* or *TeachLivE/Mursion*, have been used to create settings within which teachers can enact face-to-face teaching practices and receive coaching (e.g., Brown, Davis, & Kulm, 2011; Muir, 2012;

Dieker et al., 2017). These studies have shown improvements in teacher questioning practices and awareness of student ideas. In addition, studies have examined the impact of providing live coaching to teachers during their classroom instruction via a wireless ear-piece. In these 'bug-in-the-ear' studies, an expert outside the classroom who watches and listens to the teacher as they talked with the students can offer advice during the authentic face-to-face interactions (Wake et al., 2017; Rock et al., 2009). Teachers in these studies valued the immediacy of the feedback they were receiving and the ability for the expert outside the classroom to be less intrusive than in a traditional classroom observation.

Few studies have examined the role of natural language processing (NLP) technology for scaffolding the development and transfer of teacher responding skills. Responding to student explanations with natural NLP technology has been studied in personalized learning and intelligent tutoring contexts (e.g., Aleven & Koedinger, 2002; Walker, Rummel, & Koedinger, 2011), and automated responses have been shown to be as effective as teacher responses in many cases (e.g., VanLehn, 2011; Gerard et al., 2015). While this approach has mainly focused on responding directly to students there is an emerging interest in including the teacher in the process. For example, in addition to automatically responding to students based on an analysis of their written explanations, Gerard and Linn (2016) found that alerting teachers to give additional guidance to those students most in need was more effective than the automated responses alone. NLP technology has the potential to be used in out-of-the-classroom and face-to-face settings that are grounded in authentic contexts, and finding opportunities to leverage NLP technology to support the development and transfer of teacher responding skills is still needed.

3. The Teacher Responding Tool (TRT)

Rather than responding directly to students, the TRT is intended to bring teachers into the loop by using NLP technology to provide responding recommendations to teachers. The TRT works with students' open-response explanations of their mathematical ideas. Fig. 1 shows a screenshot of the TRT interface. At the top of the screen is the question prompt and below this is the student explanation (left, in blue) and three recommendations based on the student explanation (right). If a teacher clicks on a recommendation, the text of that recommendation is copied to the teacher response field (bottom left). Several recommendations can be clicked and added to the teacher response field. Any text in the teacher response field can be edited so that teachers can customize a recommendation or take parts of different recommendations. Recommendations can also be used without making any changes or teachers can ignore the recommendations and write their own response.

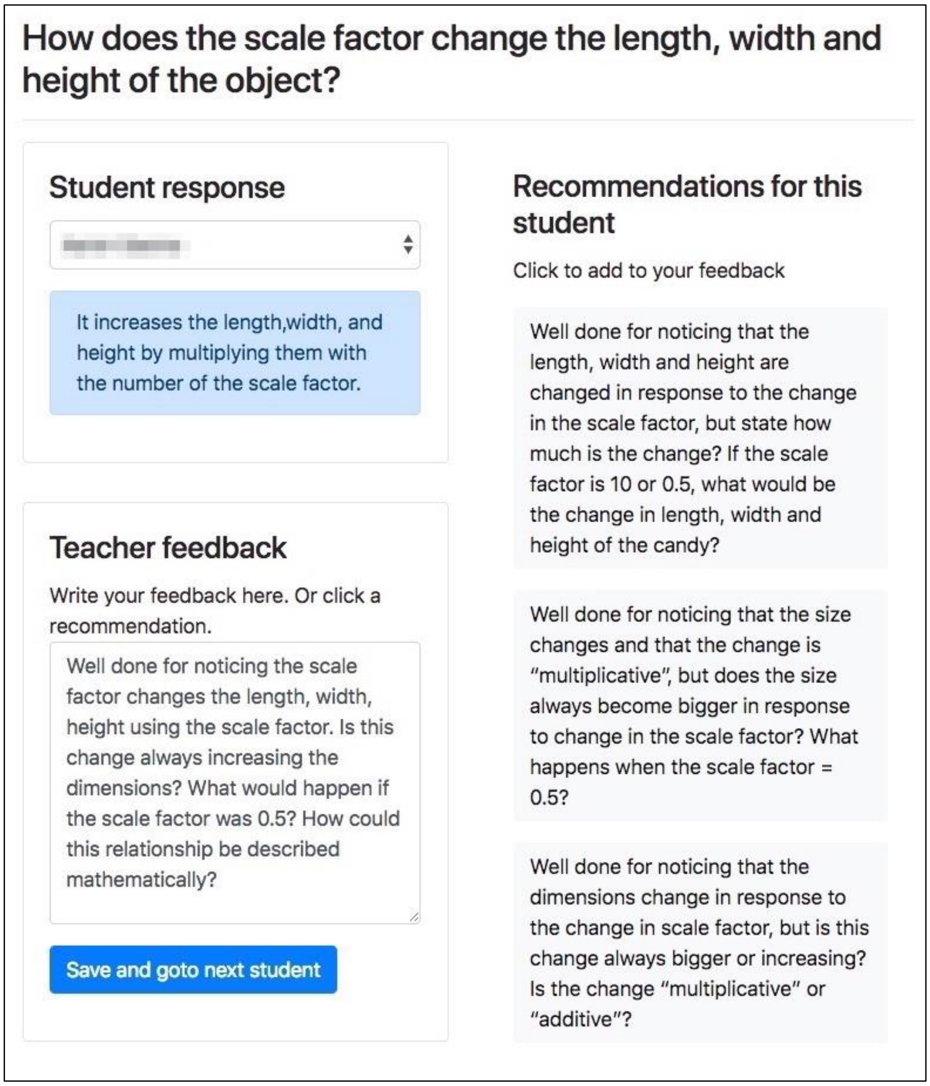


Fig. 1. Screenshots of the Teacher Responding Tool (TRT). The TRT displays the question prompt (top), the student explanation (top left, in blue), the teacher response (bottom left), and three recommendations based on the student explanation (right).

In this study, the responding recommendations were selected from a training dataset that was generated from a prior implementation of the project at the same school. During this prior implementation, a different group of students wrote explanations for the same question prompts. Afterwards, the teachers who participated in the prior implementation met for about two hours to discuss the student explanations and share ideas about how they would respond. Members of the research team used both the teacher discussion and the feedback and questioning practices described in sections 2.2 and 2.3 to craft responses for each prior student explanation. These responses highlighted a critical mathematical idea that was noticed within a student's explanation, and then asked questions to advance a student's thinking toward the intended learning goal. Table 1 is a sample of the training data generated from this process. A separate training dataset was created for the student explanations and responding recommendations for each of the three open-response questions that used the TRT.

Table 1

A sample of the 116 items in the training dataset for the open-response question: "How does the scale factor change the length, width and height of the object?"

Student Explanation	Responding Recommendation
It gets bigger	Well done for noticing that there is a change, but what is "it"? Is the change always larger? What happens when the scale factor is 10 or 0.5?
It makes the size bigger. the scale factor times by the scale	Well done for noticing that the dimensions change in response to the change in scale factor and that change is "multiplicative", but is this change always bigger or increasing?
It multiplies the original length by the number of scale factor.	Well done for noticing that the length changes in response to the change in the scale factor and that the change is "multiplicative", but what about the other dimensions – width and height?
Length and width and height multiplied by the 3	Well done for noticing that the dimensions changes and that the change is "multiplicative", but what change in the scale factor causes these changes? What happens to the dimensions when the scale factor is 0.5?
The bigger the scale factor the bigger the dimensions get.	Well done for noticing that the dimensions become bigger as the scale factor becomes bigger, but what happens when the scale factor is made smaller rather than bigger?
The new measurements are way higher than the old measurements, its shape has increased.	Well done for noticing that the size changes, but does the size always increase in response to change in the scale factor? What happens when the scale factor = 0.5? What is "it"? Expand on your answer.

For each new student explanation generated during this current study, the TRT selected three responding recommendations from the training data set. These recommendations were selected using a natural language processing (NLP) algorithm that compared how similar a new student explanation was to each of the student explanations in the training dataset. The training data set was then sorted so that the student explanations that were most similar to the new student explanation were listed first and the least similar last. The accompanying prewritten responding recommendations for the most similar responses were then examined and the top three unique recommendations were selected for the teacher to consider.

The degree of similarity between a new student explanation and an explanation in the training dataset was determined using cosine similarity—i.e. the cosine of the angle between the vectors representing the explanations. The features of these vectors were the stems of the individual words in the responses (lemmatized unigrams) and the weights were determined using tf-idf (term frequency-inverse document frequency). This approach gives higher similarity scores when comparing explanations that have the same words or word stems, but also assigns higher weights to words that occur less often in the training dataset (Zehner, Sälzer, & Goldhammer, 2016).

To assess the performance of the TRT algorithm for selecting the responding recommendations we performed a leave-one-out cross-validation with each of the training datasets (Borra & Di Ciaccio, 2010). This approach removed one student explanation from the training dataset and used the remaining training data to select recommendations for the removed student explanation. The selected recommendations were then examined to see if they contain the actual recommendation associated with the removed student explanation. If so, this is counted as a success; if not, a failure. This process was repeated, leaving out a different student explanation from the training dataset each time, until all the student explanations in the dataset had had their turn to be left out. Dividing the total number of successes by the size of the dataset gave the proportion of successes for each training dataset (see Table 2).

To account for successes that are expected by random chance we calculated kappa for each dataset (see Table 2). A kappa of 0 indicates that all the success is due to randomness and a kappa of 1 indicates success every time (Cohen, 1968). For studies that scored similar open-response items, Liu and colleagues (2014) report average kappa values between 0.62 and 0.81. While our kappa value for the question 1 dataset is lower than this, we expect this is likely due to the high number of unique recommendations included in this dataset. The studies that Liu and colleagues (2014) report on use between two and five categories, making them comparable to the question 2 and question 3 datasets using in this study. Since the kappa values for these questions are similar to those reported by Liu and colleagues (2014), we conclude that the performance of the TRT recommendation selection algorithm used in this study (see Table 2) is comparable to those used by other studies.

Table 2
Number of explanations, unique recommendations, proportion of successful recommendation selections, and kappa values by question

Training dataset	Number of student explanations	Number of unique recommendations	Proportion of successful selections	Kappa
Question 1	116	29	0.560	0.509
Question 2	99	6	0.919	0.838
Question 3	85	6	0.882	0.764

4. Method

4.1. Participants

The participants in this study were four high school geometry teachers from a small city in the rural mid-Atlantic region of the United States. The demographics of the school are 12%

Black, 44% Hispanic, and 38% White students, with 68% of the students receiving free or reduced lunch and 39% of the students classified as having Limited English Proficiency. All four teachers were white, two of the teachers were female, two were male, and their average full-time classroom teaching experience was approximately six years (see Table 3). The geometry classes that participated in this study enrolled students that the school considered to be ‘low performing’ and the classes met every other day for 88 minutes.

Table 3
Summary of participant teaching experience

Name (pseudonym)	Number of years teaching	Gender	Number of students (classes)
Sam	8	Female	20 (2)
Nina	2	Female	20 (2)
Mike	11	Male	20 (2)
Henry	4	Male	8 (1)

4.2. Instructional context: Optimizing candy packaging

The Teacher Responding Tool (TRT) was used within an engineering design project that focused on the modeling and optimization of candy packaging. The goal of the project was to redesign candy packaging to use as little paper as possible. The project was co-designed with the participating teachers and researchers to help students develop their understanding of how the surface area and volume of a prism change with scale factor. This mathematical content was selected by the participating teachers as an area of need for their students.

To support students through the project, the students used WISE (see Appendix A), which provided a rich set of features designed to support inquiry instruction. In this project, students used WISE to manipulate interactive GeoGebra visualizations, to record their measurements, calculations, and thinking, and to share their designs with other students (see Fig. 2). Students were also provided with paper, scissors, tape, ruler, and a packet of candies which they used as hands-on manipulatives (see Fig. 3). In addition, the students were encouraged to discuss their ideas with each other and to seek help from their teacher if needed.

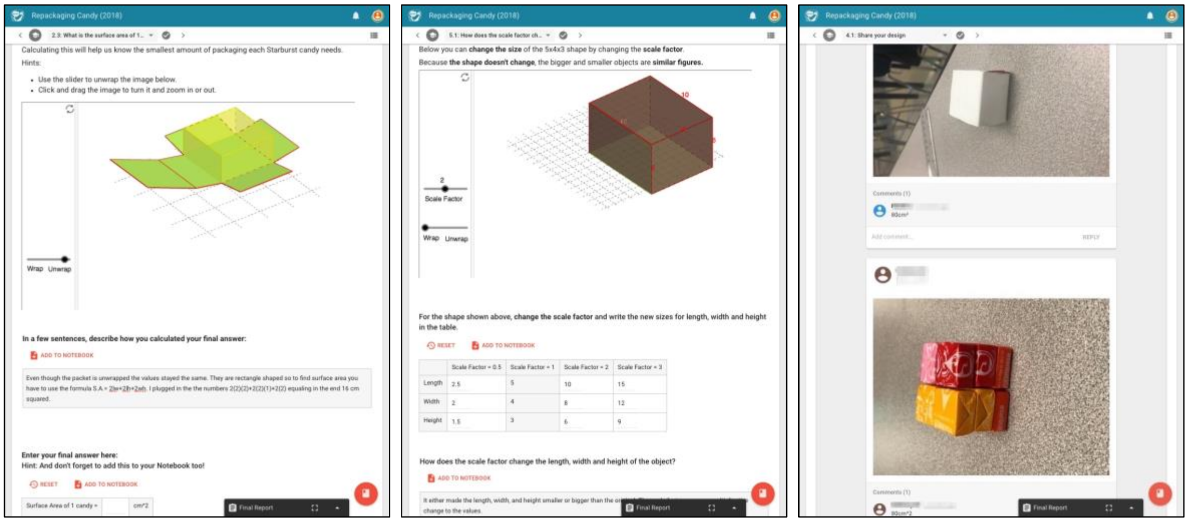


Fig. 2. Three screenshots of the project's online WISE resources.

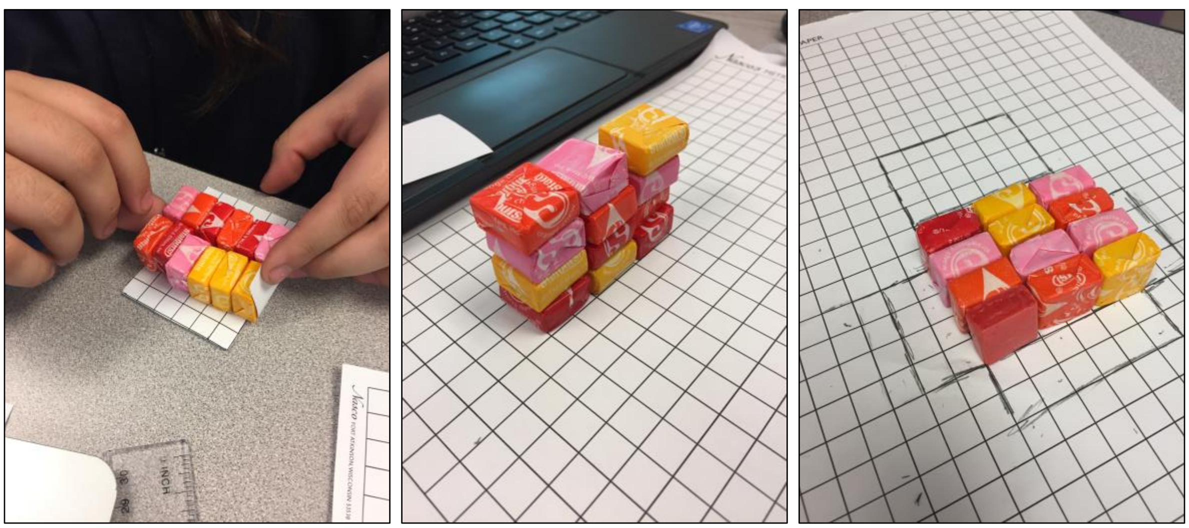


Fig. 3. Students were given hands-on manipulatives that included one packet of twelve starburst candies.

4.3. Procedure

The Optimizing Candy Packaging project ran for three class periods. Apart from introductions at the start of the project and for brief instructions at the start and end of each block, there was no whole-class direct instruction. Instead teachers moved from student to student, asking questions, listening to student explanations, making references to the online resources, giving encouragement, and providing explanations to students as needed.

During the first day of the project, the students explored the different ways that candies could be arranged in order to reduce the surface area of the outer wrapper. Their design challenge was to minimize the candy surface area while keeping the candy volume constant. To do this, the students found the surface area and volume of different candy arrangements, looked for patterns in their results, and tested which candy arrangement best met the design challenge. During the second day of the project, the students explored and modeled how changing the scale factor impacted the size, surface area and volume of a candy in the shape of a rectangular prism.

Throughout the project the students were asked to write explanations of their mathematical thinking. The question prompts for the students are shown in Table 4. After each day, the teachers responded to the student explanations that were written that day. After the first day, the teachers used the TRT interface *without* the feedback recommendations, and after the second day, the teachers used the TRT interface *with* the feedback recommendations. In both cases, the students were alerted to their teacher responses at the start of class on the following day (i.e. at the start of class on the second and the third days) and were encouraged to revise their explanations before continuing with the project.

Table 4
Question prompts used to elicit student explanations

Day 1: Teachers respond *without* recommendations	Day 2: Teachers respond *with* recommendations
How does your design meet the design challenge?	How does the scale factor change the length, width and height of the object?
What math did you use to help you meet the design challenge?	How does the scale factor change the volume of the object?
	How does the scale factor change the surface area of the object?

4.4. Data Sources

4.4.1. Teacher think-aloud.

Each teacher individually participated in a think-aloud session during which they used the TRT to read their students' open-response explanations and write responses. During a session the teacher was asked to talk about what they were thinking about as they wrote. If the teacher didn't say anything for more than about 30 seconds, they were asked by the researcher, "What are you thinking?" (Russo, Johnson, & Stephens, 1989). Throughout a think-aloud session the following data was collected: (1) an audio recording of what the teacher said; (2) researcher field notes about context and non-verbal reactions; (3) a video recording of the teacher's computer screen to capture the teacher's interactions with the TRT; and (4) the final response written by the teacher for each student explanation.

4.4.2. Teacher post-project interview.

After the final day of the project, each teacher participated in an interview that was audio recorded and lasted on average 40 minutes. This semi-structured interview focused on the teacher's responding practices, and the impact, if any, of using the TRT. All teachers were asked the same questions (see Appendix B) but additional probing questions or modifications to the questions were made in order to collect richer teacher descriptions or to maintain the flow of the interview.

4.5. Data Analysis

4.5.1. Length of teacher responses and responding time.

The average length of the teacher responses was found by counting the number of characters in the responses and dividing by the number of responses. The average teacher responding time was found by dividing the duration of a teacher's think-aloud sessions by the number of responses written.

4.5.2. Teacher selection of the TRT recommendations.

From the video recording of the teachers' computer screens, we recorded which recommendations were used, if any, and whether teachers chose to edit or combine recommendations, or not.

4.5.3. Degree of responding to student ideas.

Building upon a rubric used by Jacobs and colleagues (2010), we developed a rubric to assess the degree to which the teacher response contained evidence of responding to mathematical ideas in the student explanation (see Table 5). Twenty percent of the teacher responses were randomly selected and coded independently by two researchers. The codebook was revised until greater than 90% agreement was obtained, after which the rest were scored by one researcher.

Table 5

Rubric used to code the teacher responses for the degree of responding to student ideas

Code	Criteria	Example [a]
2 – *Robust* evidence that teachers were responding to student ideas.	The teacher response is consistent with the student explanation, **and** it makes a specific interpretation of what the student <u>does</u> understand (not general comments such as "I like your idea" or "Your thinking is great") **and** it makes a specific interpretation of what the student ought to consider next (what to think about, what to clarify etc.) either as statements or questions.	*Student explanation:* "It is multiplied by 4 each time" *Teacher response:* "Good job finding the multiplicative relationship with a scale factor of 2. What happens at a scale factor of 3? Can you find a relationship between the different scale factors and their associated surface areas?"
1 - *Limited* evidence that teachers were responding to student ideas.	The teacher response is consistent with the student explanation, **but** it only asks for clarification or elaboration, **or** it only states what the student has not yet articulated.	*Student explanation:* "It is multiplied by 4 each time" *Teacher response:* "What is multiplied by 4? You have not yet said how scale factor changes the surface area"

	The teacher response can involve restating or rewording of the question if consistent with the student explanation.	
0 – *Lack of* evidence that teachers were responding to student ideas.	The teacher response may include specific interpretations but these are inconsistent with the student explanation, **or** the teacher response is general and doesn't indicate an interpretation of student ideas, **or** the teacher response is only generic praise (e.g. "well done" or "keep going").	*Student explanation:* "It is multiplied by 4 each time" *Teacher response:* "This is great so far, but you haven't fully answered the question."
No score	If student explanation is blank or off task there is nothing to notice or interpret.	*Student explanation:* "…"

[a] These examples relate to the question prompt: "How does the scale factor change the surface area of the object?"

4.5.4. Teacher description of the impact of the TRT on their practice.

The transcripts of the teacher post-project interviews and the teacher think-aloud sessions were analyzed to identify reoccurring themes and reanalyzed to identify evidence that confirmed or diverged from the themes.

5. Results

5.1. How do teachers interact with the TRT recommendations?

In total, the teachers responded to 82 student explanations using the TRT recommendations—not all students wrote explanations to all the questions. Across all teachers, approximately three-fourths of the teacher responses made direct use of the recommendations with half (41) of the teacher responses using a recommendation and editing it, and about one fourth (23) of the teacher responses using a recommendation without editing it. The remaining teacher responses (18) were written without the teacher selecting a recommendation, but the wording of these responses was often similar to one of the recommendations (see Table 6).

Table 6

Average responding time (in seconds), the average response length (in characters), and the use of recommendations in responding, by teacher.

Teacher	Average responding time (seconds)	Average response length (characters)	Number of responses *without* recommend-ations	Number of responses *with edited* recommend-ations	Number of responses *with unedited* recommend-ations	n
Sam	76.7	154.4	5 (56%)	3 (33%)	1 (11%)	9
Nina	69.6	185.2	4 (14%)	21 (75%)	3 (11%)	28
Mike	48.2	182.0	4 (14%)	7 (25%)	17 (61%)	28
Henry	84.7	233.1	5 (29%)	10 (59%)	2 (12%)	17
Total	66.2	190.7	18 (22%)	41 (50%)	23 (28%)	82

There were also differences between the teachers in how they interacted with the TRT recommendations. Sam wrote her responses without selecting the recommendations more often than the other teachers but also tended to write shorter length responses. Nina edited recommendations more than the other teachers, but the final response was an average length and she took an average amount of time to make these edits. Mike selected recommendations and did not edit them more often than the other teachers, and his responding time was faster than the other teachers. Henry selected recommendations and edited them at an average rate, but his responses were much longer in length than average, and they took more time to write (see Table 6).

Comparing the average responding times when the teachers did and did not use the TRT interface, we find that the time spent per response decreased from about 1.5 minutes without the recommendations to 1 minute with the recommendations (see Table 7). However, this change was largely due to one teacher and the overall picture is mixed; two of the teachers' times increased and one stayed roughly the same. Comparing the average length of the teacher responses when the teachers used the TRT interface, we find that the length per response increased when the recommendations were provided, with two teachers (Sam and Henry) having large increases to the length of their responses. This result is not surprising given that the text from the recommendations could be easily copied into the teacher response field. Of particular interest is that the correlations between responding time and response length is strong (r = 0.96) without the recommendations, but weak (r = 0.37) with the recommendations (see Table 7).

Table 7

Average responding time (in seconds), the average response length (in characters), for teacher use of the TRT with and without recommendations

	With recommendations			Without recommendations		
Teacher	Average responding time (seconds)	Average response length (characters)	n	Average responding time (seconds)	Average response length (characters)	n
Sam	76.7	154.4	9	50.3	99.7	17
Nina	69.6	185.2	28	76.7	164.6	26
Mike	48.2	182.0	28	154.0	218.1	30
Henry	84.7	233.1	17	50.0	112.1	12
Total	66.2	191.7	82	94.9	163.1	85

The teachers were able to intuitively interact with the TRT interface. None of the teachers asked questions about how to use the interface or expressed frustrations with the interface while responding. Instead teachers spent their time reading the recommendations, considering the student explanations, and writing or editing their response. For example, Mike described that when he interacted with the TRT he might say, "'Oh, okay. That's the response I want,' [and] click on it" but that at other times he would "look at the recommendations and think, 'Well, that one clearly isn't what I see happening here. This one is the closest to [the student explanation], but I think I need to just qualify it a little bit, modify it to fit this situation.'" The thoughtfulness of the teacher interactions was also apparent. It may have been the case that the teachers quickly clicked the top recommendation for each student explanation and responded with minimal consideration of the mathematical ideas. But this was not observed in this study. Instead, the teachers considered the merits of the recommendations with respect to the student explanation and interacted with the TRT based on this thinking. Nina said that it was an advantage that the recommendations were not “everything I want to say as verbatim…because if it was exactly like what I wanted to say, then I feel like that would just be a little more mindless for me…and I would say, ‘Yup. This is good enough.’ But because none of them were spot on, or they all kind of touched on different aspects, that it was easier for me to piece them together, to edit them, to change a word here or there to personalize it.” The teacher interactions with the TRT recommendations were thoughtful and focused on the responding task.

5.2. To what degree do the teacher responses relate to their students’ ideas?

Overall, we found greater evidence of teachers responding to student ideas when using the TRT recommendations. The responses indicate that, when using the TRT recommendations, the teachers were able to notice and interpret specific mathematical ideas within the student explanations, and make appropriate statements or questions for the students to consider next (see Table 8). The degree to which the teacher responses related to their students’ ideas was significantly higher when the teachers responded *with* rather than *without* the TRT recommendations ($t = 6.41$; $p < 0.000$) with a large effect size ($g = 0.99$). This increase was true for all teachers involved in the study (see Table 8).

Table 8
Average scores for the degree of teacher responding to student ideas without and with the TGT recommendations along with Welch’s t-test results.

	Without recommendations		With recommendations		Difference (with-without)			
Teacher	Mean	n	Mean	n	t	df	p	Hedges’ g
Sam	1.18	17	1.78	9	3.08	19.24	0.006	1.20
Nina	1.63	24	2.00	26	3.72	23.00	0.001	1.20
Mike	1.69	29	1.96	24	2.77	39.67	0.008	0.72
Henry	1.42	12	1.94	17	2.60	13.06	0.022	1.13
Total	1.52	82	1.95	76	6.42	109.06	0.000	0.99

[a] Note that the number of teacher responses indicated in this column may be lower than in Table 6 and Table 7 because no score was assigned when a student explanation showed no mathematical ideas (see Table 5).

5.3. How do the teachers describe the impact of the TRT on their responding practice?

The teachers described the recommendations as scaffolding their responding practice while using the TRT. Henry commented that he would sometimes look at a student response and think “Oh, goodness, where do I even begin?” but that “it was nice to have [the recommendations say], ‘Well, how about begin here?’” Mike “liked having the three and then picking one of those three...to help give some vocabulary” because sometimes when responding to students he would find himself “a little bit of a loss for words” and “respond very similarly to everybody.” And Sam, knowing that the recommendations were based on a prior teacher discussion, commented that the TRT “was helpful to see how other people approach their students.” The recommendations seemed to scaffold responding by providing teachers with ideas to help them get started and making responding a less isolating practice.

The teachers described that the recommendations helped them recognize nuances in the student explanations that they might not have otherwise noticed. Rather than seeing just “one thing,” in a student response without the recommendations, Sam explained that because the recommendations were “more specific” they gave her “other suggestions” for what to look for. For Henry, the recommendations helped his noticing of student ideas because they “gave a good sense of what kind of answers to look for, as to whether or not they mentioned that ‘the one just affected the other,’ or the specific way they said ‘as the one increased, the other increased,’ or if they then went so far as to recognize that there was a multiplicative relationship.” Without the recommendations, Nina commented that after reading what her students had written, she “categorized all their responses…and said the same thing just about for each of them”, but that with the recommendations she was able to notice “the subtle differences” in the mathematical ideas of her student explanations. This led Nina to comment that the TRT recommendations "gave me way more information and way more direction, and prepared me so much better for the next class."

The teachers also said that the recommendations helped them offer guidance about the next steps a student might think about based on their current understanding rather than evaluating the current understanding of a student. Henry commented that he would typically "identify exactly what's wrong in an individual problem and then, hope that they use that knowledge going forth with the further problems." But he said that with the recommendations, he was better able to "give a tailored response to give them a hint towards where they supposed to be going" and that "this is not necessarily a straightforward task because there were plenty of other future steps for them to potentially make." Nina also commented that the TRT recommendations "kind of enabled [her] to focus on those subtleties and draw out information from [students]" by asking questions such as "What do you mean by a lot? What is the relationship there?"

The teachers also commented on how the recommendations were much more positive than how they would typically respond. Nina pointed out that this was something she would "always forget, especially when typing into a computer instead of talking face to face with a person." Sam remarked that giving positive comments was "just not my style" and Henry commended that he was typically "much more direct" with his responses to students. That said, the teachers didn't remove any of the positivity from the recommendations they selected and appreciated this aspect of them. Mike stated that he normally needs to remind himself "to Oreo cookie it…positive, correction, positive," and that "a lot of the students don't necessarily appreciate" feedback without the sandwiching. The positivity in the recommendations helped to remind him of this strategy and so he "appreciated having some of those words there." And for

Nina, it made “a big difference” that the TRT encouraged her to respond more positively because it is “really important, [in her] own practice.”

Some of the teachers also remarked that using the TRT gave them opportunities to reflect on their responding practice. For Nina, reflecting on teaching practice “was a huge part of [her] pre-service teacher curriculum” but was lacking in her day-to-day work schedule despite it being “as beneficial as it is.” She described the use of the TRT not as extra work or grading, but as an opportunity to reflect. She said that she “really liked setting aside time” for thinking about how best to respond to students.

6. Discussion

In this study we conjectured that providing automated recommendations to teachers with the TRT can help teachers develop their practice of responding to student ideas, or more specifically, develop their practice of noticing student ideas, providing formative feedback to these ideas, and asking questions to develop student thinking. To examine this conjecture, we asked research questions about the role of the TRT recommendations on the observable interactions of the teachers with the tool (research question 1), on the responding artifacts produced by teachers (research question 2), and on the teachers’ perceptions about the development of their responding practice (research question 3). This discussion will focus on each of these mediating processes and their impact on our conjecture (Sandoval, 2014).

Our results indicate that when provided with recommendations, the teacher interactions with the TRT involved the teachers reading a student explanation and choosing and editing recommendations, or not, based on thoughtfully considering the recommendations alongside the student explanation. In contrast, when responding without the recommendations, the teachers reported seeing less nuance in the student explanations. The three TRT recommendations appear to have made the practice of responding more complex because they gave teachers additional ideas to think about alongside the student explanation.

At the same time, teachers reported that the recommendations made responding easier because the tool presented choices for teachers to pick and edit rather than a blank space. What may sound like a paradox—that the TRT recommendations make responding both more complex and easier at the same time—indicates that the interactions with the tool may have changed the processes teachers use to respond to student explanations. The TRT problematized the responding task for the teachers by highlighting the subtleties in student mathematical ideas and the complexities involved in noticing. The teacher interactions with the TRT did not simplify the complex task of responding to a simple mouse-click exercise. Such a simplification may have made responding easier for teachers but would have done little to develop their practice. In this sense, the TRT aligns well with other tools that are designed to scaffold complex learning (e.g., Reiser, 2004).

The teacher responses with the TRT recommendations demonstrated a high degree of responding to student ideas. The degree of teachers' responding to student ideas with the TRT recommendations was significantly higher than without the TRT recommendations. To some degree the difference could be attributed to the TRT algorithm automatically presenting recommendations to teachers that were similar to the ideas in the student explanation. However, the differences are unlikely to have been observed without teachers noticing and responding to the student ideas because not all the recommendations presented by the TRT would have aligned closely enough with the student explanations to score as highly as was observed.

Despite the observed improvement in responding to student ideas, the time that the teachers spent responding was not clearly different, with some teachers spending more and others less time. The reason for this appears to be in how the time was spent. Without the recommendations, there was a strong positive relationship between the time a teacher spent responding and the length of their responses, suggesting that without the recommendations much of the time that the teachers spent responding was spent typing. This relationship was weak when the recommendations were provided.

In addition to improving responding to student ideas, the interviews revealed that the positive, strength-based format of the TRT recommendations was appreciated by teachers. The recommendations were written to value the ideas that students currently held and asked questions to advance student thinking. Teachers reported that this was different from their typical practice which tended to be more evaluative and focus on what students had not explained or what was incomplete about their explanations. Given the importance of strength-based approaches and similar growth mindsets in mathematics (e.g., Haimovitz & Dweck, 2017), teacher-in-the-loop tools such as the TRT may help model and guide other important practices.

7. Implications

7.1. Implications for teaching and learning

While there has been much attention on how to leverage machine learning technology to support student learning, results suggest that such technologies can also be used to support teacher learning and teacher reflection upon student ideas. Although it was not examined in this study, the TRT offers an opportunity to support the development of teachers' responding practice with students in a face-to-face classroom setting. Future work will explore to what extent teachers' responding practices change during subsequent classroom interactions after working with the TRT, and to what extent teachers' responding practices have on student learning.

Our results highlight that the having the TRT recommendations alongside the student explanation made the nuances in student thinking more visible than with the student explanation alone. In addition to making teachers' thinking visible to students (e.g., Collins, Brown, & Holum, 1991; Selling, 2016), and providing multiple representations of concepts to help students understand content, classrooms need to provide opportunities for students to express their understanding in different ways. Noticing and responding to student ideas requires teachers to have access to representations of students' ideas. This study implies that it is important to include tasks and activities in mathematics classrooms that elicit student thinking but that it is also important to use tools that help make students' thinking more visible to teachers.

7.2. Implications for the design of responding tools

The TRT was designed so that teachers were presented with three recommendations from which to choose. This intentionally problematized the practice of responding for teachers, requiring the differences between recommendations to be considered alongside the student explanation. This design is different from how natural language processing (NLP) is typically used within responding tools. Typically, such tools are used to auto-respond for the teacher, allowing teachers to focus their time where most needed (Gerard et al., 2015). However, such approaches can limit the visibility of student thinking to teachers. The TRT design addresses this issue by positioning the teacher as the responding decision maker, asking them to make choices but also allowing them to edit or personalize their choices.

The recommendations were written to follow evidence-based formative feedback practices. This included targeted domain-specific praise of what students did understand and questions about what students might next think about. Teachers reported that this strength-based format positively impacted their responses. However, the kinds of data presented to teachers in computer-based learning environments typically focus on scoring responses for correctness, highlighting errors, or tracking progress (e.g. Xhakaj, Aleven, & McLaren, 2017). Thus, results from this study highlight the need and potential of other computer-based environments to model strength-based approaches for teachers in mathematics contexts.

Given that the TRT presents recommendations to teachers about how to respond, the importance of highly accurate NLP algorithms is mitigated. High accuracy is often achieved by constraining the questions asked, by the generalizing the responses given, or by using large training datasets. By providing three recommendations filtered through the teacher, we found that these constraints can be relaxed somewhat and more open questions can be asked, more specific responses can be provided, and smaller training datasets can be used. Future research with natural language processing technologies can explore similar benefits of efficiency and accuracy with teacher-in-the-loop approaches.

While this study has focused on a providing responding recommendations to teachers within the domain of mathematics, the tool is not limited to this domain. The TRT interface currently has the potential be applied broadly to any setting where an instructor could be supported by formative feedback recommendations for written learner responses in computer-based environments. For example, the TRT could be used in science and social science classrooms, in workplace training contexts, and with non-English responses. Furthermore, since other kinds of student-generated data such as graphs, charts, or tables can be represented as text, the tool could be trained to provide response recommendations for these other kinds of student generated data. Future work can explore the potential and generalizability of the TRT in different domain contexts.

7.3. Implications for teacher professional development

The TRT has the potential to fit within a professional development "cycle of enactment and investigation" (Lampert et al., 2013) that supports teachers in authentic classrooms integrated with everyday tasks. Instead of typical professional development models that are highly structured, time and resource intensive, this study highlights the potential of providing supplementary teacher learning opportunities with focused technologies dedicated to help teachers with a specific practice. This kind of just-in-time, highly contextualized and targeted approach may fill a need to support teachers to develop, reflect upon, and refine their practice. Future work can also investigate how these opportunities can complement more traditional professional development models.

8. Limitations

One limitation of the scope of study is that we did not follow teachers responding practice into the classroom and observe how their responding while using the TRT may or may not have transferred into the classroom. That said, the teachers did indicate that having used the TRT they felt better prepared for talking with their students in class the next day because they understood the kinds of mathematical ideas that their students held. Further research into the impact of the TRT on classroom face-to-face responding is still needed, but we think that our results present a promising new approach to scaffolding the teaching practice of responding to student ideas.

This study has other limitations such as our ability to generalize these specific results, that were found from studying four teachers in one school during one instructional context. In addition, while we did make comparisons between two conditions—responding with and without the recommendations—different question prompts were asked in each case, and the observed differences in responding to student mathematical ideas may have been as a result of differences in the teachers' mathematics knowledge for teaching (MKT; Ball, Hill, & Bass, 2005) rather than because of the TRT recommendations.

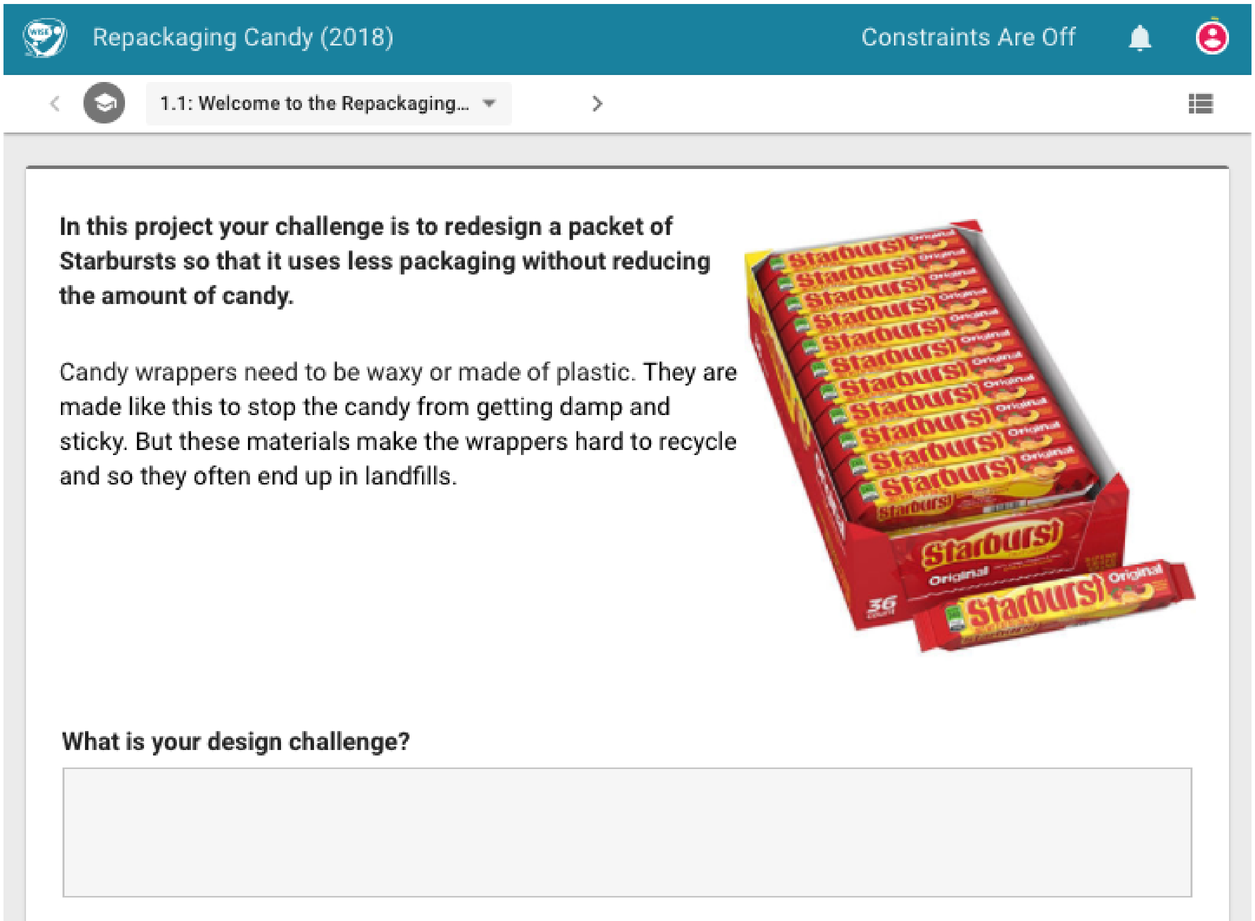
The tool itself also has important limitations to consider. In focusing on students' text-based explanations, the TRT does not support teacher responding in contexts where students may be better able to draw or speak their ideas. In addition, unless the TRT is trained with multiple languages, it will be less effective in classrooms with linguistic diversity. Further research might consider the use of voice-to-text applications which allow students to speak their ideas rather than type them, or might examine extending the tool to incorporate other forms of student explanations such as diagrams or equations.

9. Conclusions

Research in teacher education emphasizes the need to provide multiple opportunities for teachers to develop high-leverage teaching practices (Lampert et al., 2013). This research advances the field of teacher education by leveraging natural language processing technology—typically used only in intelligent tutoring contexts—to scaffold the teacher practice of responding to student ideas. By introducing and implementing the TRT, this research examines a novel approach for teachers to develop their noticing and responding practice in real-time, authentic and meaningful classrooms contexts. The positive findings in this research indicate that these kinds of tools could potentially play a significant role in supporting teachers' ongoing development of responding practice.

Appendix A

Screenshots of the some of the instructional materials and resources accessed by students.



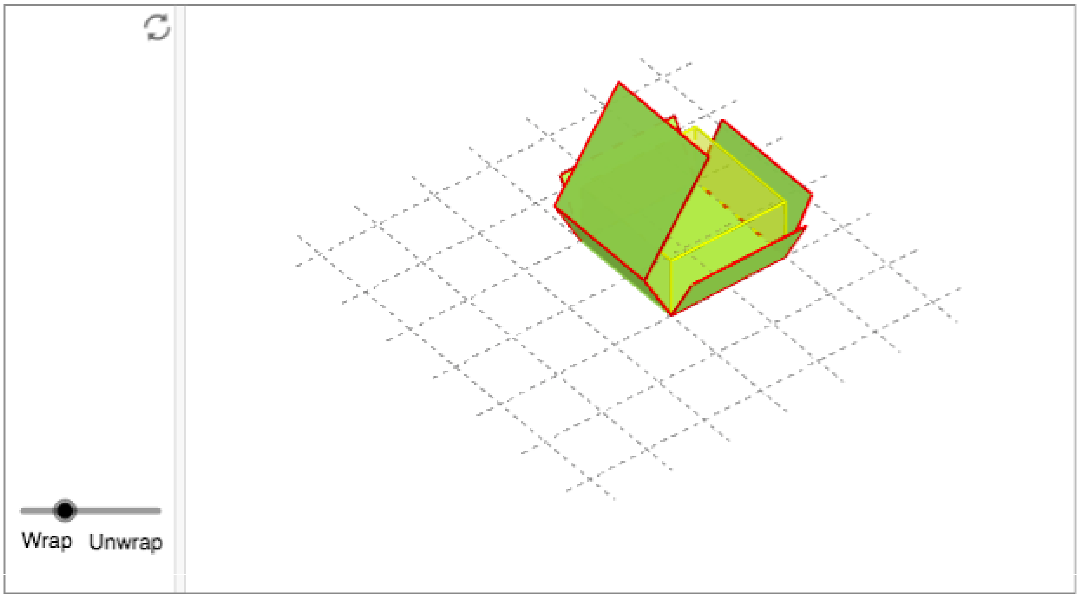
2.3: What is the surface area of 1...

What is the surface area of 1 candy?

Calculating this will help us know the smallest amount of packaging each Starburst candy needs.

Hints:

- Use the slider to unwrap the image below.
- Click and drag the image to turn it and zoom in or out.



In a few sentences, describe how you calculated your final answer:

Enter your final answer here:
Hint: And don't forget to add this to your Notebook too!

RESET ADD TO NOTEBOOK

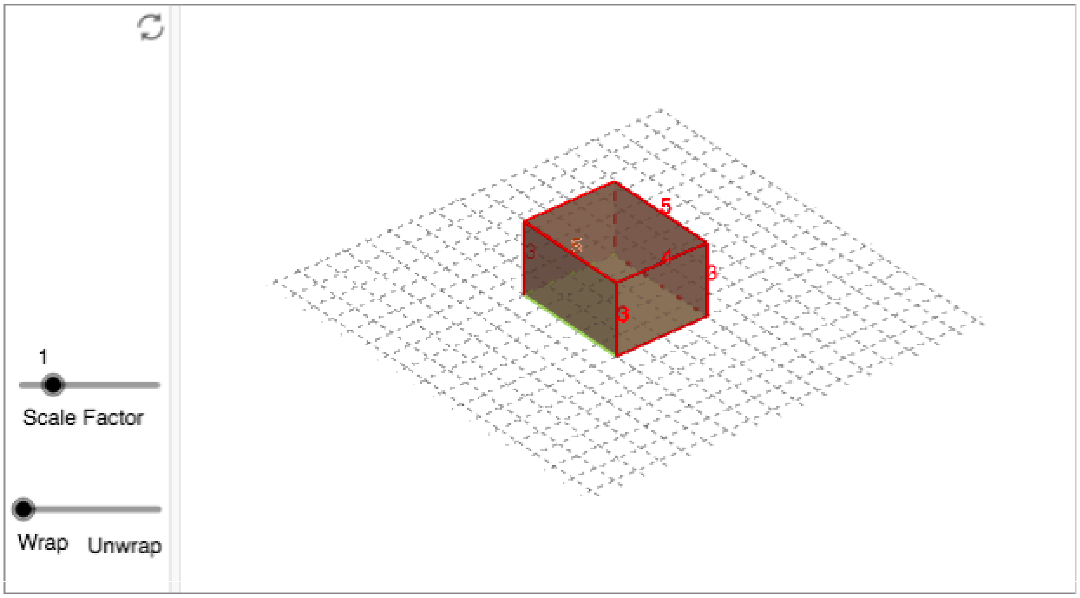
Surface Area of 1 candy =		cm^2

5.1: How does the scale factor ch...

A different shape.

Below you can **change the size** of the 5x4x3 shape by changing the **scale factor**.

Because **the shape doesn't change**, the bigger and smaller objects are **similar figures.**

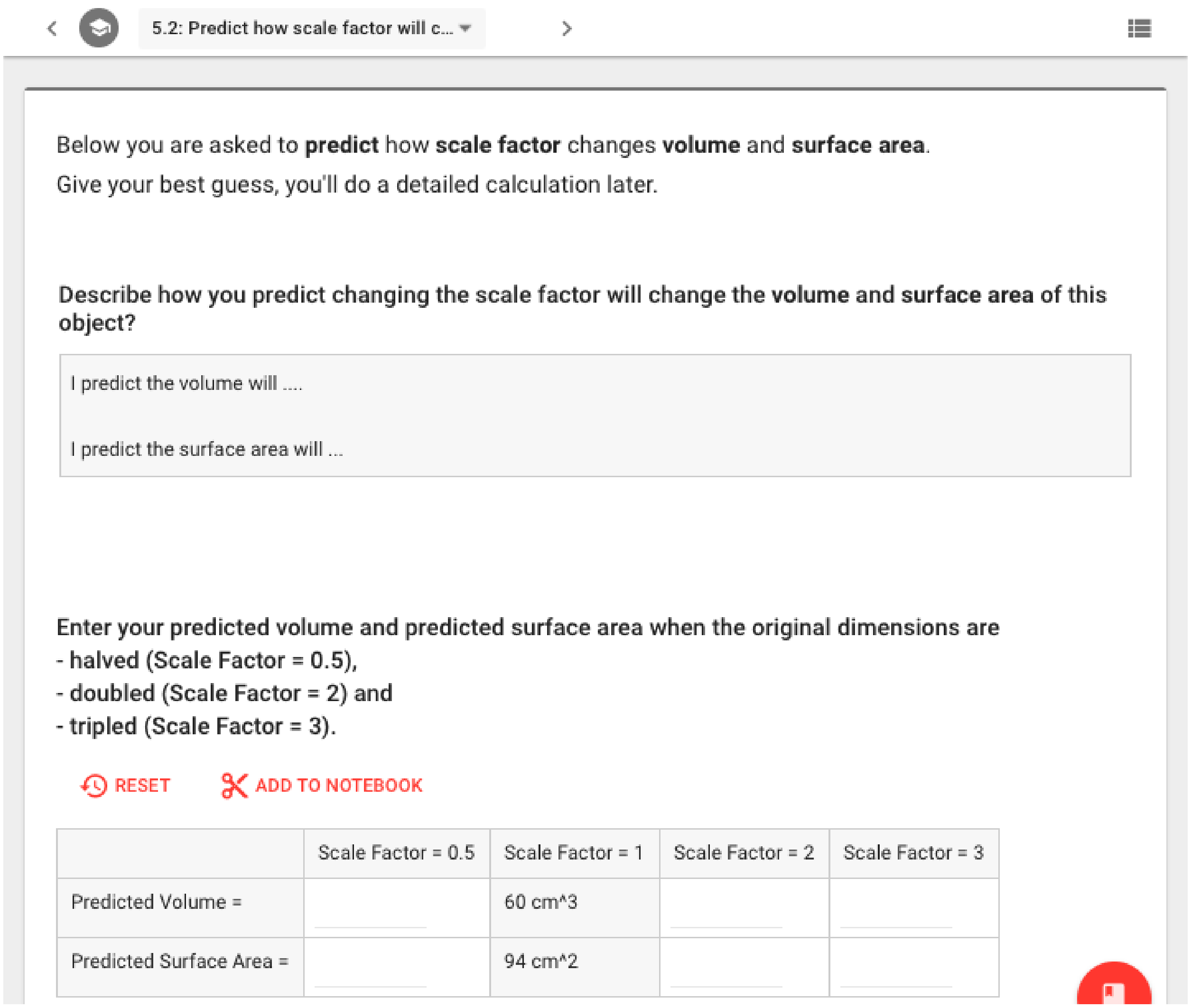


For the shape shown above, change the scale factor and write the new sizes for length, width and height in the table.

RESET ADD TO NOTEBOOK

	Scale Factor = 0.5	Scale Factor = 1	Scale Factor = 2	Scale Factor = 3
Length		5		
Width		4		
Height		3		

How does the scale factor change the length, width and height of the object?


5.2: Predict how scale factor will c...
Below you are asked to predict how scale factor changes volume and surface area.
Give your best guess, you'll do a detailed calculation later.
Describe how you predict changing the scale factor will change the volume and surface area of this object?
I predict the volume will ....
I predict the surface area will ...
Enter your predicted volume and predicted surface area when the original dimensions are
- halved (Scale Factor = 0.5),
- doubled (Scale Factor = 2) and
- tripled (Scale Factor = 3).
RESET
ADD TO NOTEBOOK
Scale Factor = 0.5
Scale Factor = 1
Scale Factor = 2
Scale Factor = 3
Predicted Volume =
60 cm^3
Predicted Surface Area =
94 cm^2

5.3: How does scale factor chang...

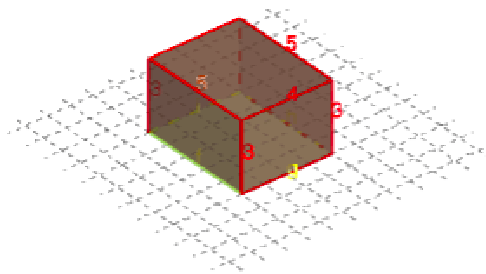
How does scale factor change the volume?

Volume
= length x width x height
= 5 x 4 x 3
= 60cm^3

1

Scale Factor

Wrap Unwrap



Adjust the Scale Factor to 0.5, 2, and 3, and calculate the volume of the object.

RESET   ADD TO NOTEBOOK

	Scale Factor = 0.5	Scale Factor = 1	Scale Factor = 2	Scale Factor = 3
Volume =		60 cm^3		

How does the scale factor change the volume of the object?

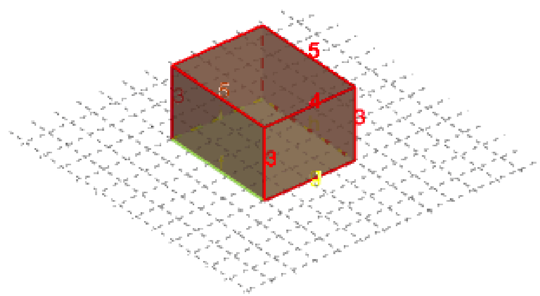
5.4: How does scale factor chang...

How does scale factor change the surface area?

Surface Area
= 2 l w + 2 w h + 2 h l
= 2(5)(4) + 2(4)(3) + 2(3)(5)
= $94 cm^2$

1
Scale Factor

Wrap Unwrap



Adjust the scale factor to 0.5, 2, and 3 and calculate the **surface area** of the object.

RESET ADD TO NOTEBOOK

	Scale Factor = 0.5	Scale Factor = 1	Scale Factor = 2	Scale Factor = 3
Surface Area =		94 cm^2		

How does the scale factor change the surface area of the object?

5.5: How does scale factor change...

Compare your predictions with your calculations and answer the question below.

Your predictions:

RESET ADD TO NOTEBOOK

	Scale Factor = 0.5	Scale Factor = 1	Scale Factor = 2	Scale Factor = 3
Predicted Volume =	.	60 cm^3		
Predicted Surface Area =		94 cm^2		

Your volume calculations:

RESET ADD TO NOTEBOOK

	Scale Factor = 0.5	Scale Factor = 1	Scale Factor = 2	Scale Factor = 3
Volume =	.	60 cm^3		

Your surface area calculations:

RESET ADD TO NOTEBOOK

	Scale Factor = 0.5	Scale Factor = 1	Scale Factor = 2	Scale Factor = 3
Surface Area =	.	94 cm^2		

In a few sentences below, describe what have you learned about how **scale factor** changes
- the dimensions,
- the surface area and
- the volume
of a 3D shape?

Appendix B

Post-project Interview Questions.

Tell me about…

- your teaching background
- how you teach
- how you give guidance
- this project
- how you gave guidance during this project in class
- how you gave guidance with the writing recommendations
- how you used the interface
- the recommendations
- how you used the recommendations?
    - Did you make changes?
    - Did you write your own?

Other thoughts?

References

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*, 147–179.

Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of school mathematics. *The Elementary School Journal*, *93*(4), 373–397.

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), 14–17, 20–22, 43–46.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice, 5*(1), 7-74.

Borra, S., & Di Ciaccio, A. (2010). Measuring the prediction error. A comparison of crossvalidation, bootstrap and covariance penalty methods. *Computational Statistics & Data Analysis, 54*, 2976-2989.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academies Press.

Brown, I. A., Davis, T. J., & Kulm, G. (2011). Pre-service teachers' knowledge for teach- ing algebra for equity in the middle grades: A preliminary report. *The Journal of Negro Education, 80*(3), 266-283.

Burns, M. (1995). *Writing in math class: A resource for grades 2-8*. Math Solutions

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (1999). *Children's mathematics: Cognitively Guided Instruction.* Portsmouth, NH: Heinemann.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school*. Portsmouth, NH: Heinemann.

Chapin, S. H., O'Connor, C., & Anderson, N. C. (2009). *Classroom discussions: Using math talk to help students learn, Grades K-6.* Sausalito, CA: Math Solutions.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, *13*(2), 145-182.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219-243.

Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching*, *48*(10), 1109–1136.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American educator*, *15*(3), 6-11.

Countryman, J. (1992). *Writing to learn mathematics: Strategies that work*. Portsmouth, NH: Heinemann.

Dieker, L. A., Hughes, C. E., Hynes, M. C., & Straub, C. (2017). Using simulated virtual environments to improve teacher performance. *School University Partnerships: Technology to Enhance PDS*, *10*(3), 62-81.

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.

Gerard, L. F., & Linn, M. C. (2016). Using Automated Scores of Student Essays to Support Teacher Guidance in Classroom Inquiry. *Journal of Science Teacher Education*, *27*(1), 111–129.

Gerard, L., Matuk, C., McElhaney, K., & Linn, M. C. (2015). Automated, adaptive guidance for K-12 education. *Educational Research Review*, *15*, 41–58.

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

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice*, *15*(2), 273–289.

Haimovitz, K., & Dweck, C. S. (2017). The origins of children's growth and fixed mindsets: New research and a new proposal. *Child development*, *88*(6), 1849-1859.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112.

Herbel-Eisenmann, B. A., Steele, M. D., & Cirillo, M. (2013). (Developing) teacher discourse moves: A framework for professional development. *Mathematics Teacher Educator, 1*(2), 181.

Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional Noticing of Children's Mathematical Thinking. *Source: Journal for Research in Mathematics Education Journal for Research in Mathematics Education*, *41*(2), 169–202.

Kazemi, E., & Stipek, D. (2009). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *Journal of Education, 189*(1-2), 123-137.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119* (2), 254–284.

Lampert, M., Franke, M., Kazemi, E., Ghousseini, H., Turrou, A., Beasley, H., & Crowe, K. (2013). Keeping it complex: Using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education, 64*(3), 226-243.

Liu, O. L., Brew, C., Blackmore, J., Gerard, L., Madhok, J., & Linn, M. C. (2014). Automated scoring of constructed-response science items: Prospects and obstacles. *Educational Measurement: Issues and Practice*, *33*(2), 19–28.

Ma, T., Brown, I. A., Kulm, G., Davis, T. J., Lewis, C. W., & Allen, G. D. (2014). Constructing and Role-Playing Student Avatars in a Simulation of Teaching Algebra for Diverse Learners. *Urban Education*, *51*(5), 534–555.

McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core Practices and Pedagogies of Teacher Education: A Call for a Common Language and Collective Activity. *Journal of Teacher Education*, *64*(5), 378–386.

McElhone, D. (2013). Pressing for elaboration in student talk about texts. *Journal of Classroom Interaction*, 4-15.

Muir, T. (2012). Virtual mathematics education: Using second life to model and reflect upon the teaching of mathematics. In J. Dindyal, L. P. Cheng, & S. F. Ng (Eds.), *Mathematics education: Expanding horizons. Proceedings of the 35th Annual Conference of the*

Mathematics Education Research Group of Australasia (pp. 521–528). Singapore: MERGA.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Ed.), *Instructional design for multimedia learning* (pp. 181–195). Munster, NY: Waxmann.

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

National Council of Teachers of Mathematics (2018). *Catalyzing change in high school mathematics: Initiating critical conversations*. Reston, VA: Author.

O'Connor, M. C., & Michaels, S. (1993). Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology & Education Quarterly*, *24*(4), 318–335.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*(3), 273–304.

Robertson, A. D., Atkins, L. J., Levin, D. M., & Richards, J. (2016). What is responsive teaching? In A. D. Robertson, R. E. Scherr, & D. Hammer (Eds.), *Responsive teaching in science and mathematics* (pp. 1– 34). New York: Routledge.

Rock, M. L., Gregg, M., Thead, B. K., Acker, S. E., Gable, R. A., & Zigmond, N. P. (2009). Can you hear me now? Evaluation of an online wireless technology to provide real-time feedback to special education teachers-in-training. *Teacher Education and Special Education*, *32*(1), 64-82.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, *17*(6), 759-769.

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the learning sciences*, *23*(1), 18-36.

Selling, S. K. (2016). Making mathematical practices explicit in urban middle and high school mathematics classrooms. *Journal for Research in Mathematics Education*, *47*(5), 505-551.

Sherin, M. G. (2002). A balancing act: Developing a discourse community in a mathematics classroom. *Journal of Mathematics Teacher Education*, *5*(3), 205–233.

Sherin, M., Jacobs, V., & Philipp, R. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York: Routledge.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189.

Slotta, J. D., & Linn, M. C. (2009). *WISE science: Web-based inquiry in the classroom*. New York, NY: Teachers College Press.

Smith, M. S., Steele, M. D., & Raith, M. L. (2017). *Taking action: Implementing effective mathematics teaching practices*.

Stoyle, K. L., & Morris, B. J. (2017). Blogging mathematics: Using technology to support mathematical explanations for learning fractions. *Computers & Education*, *111*, 114-127.

Teuscher, D., Kulinna, P. H., & Crooker, C. (2016). Writing to learn mathematics: An update. *The Mathematics Educator*, *24*(2).

van Es, E. A., & Sherin, M. G. (2002). Learning to Notice: Scaffolding New Teachers' Interpretations of Classroom Interactions. *Jl. of Technology and Teacher Education, 10*(4), 571–596.

van Es, E. A., & Sherin, M. G. (2010). The influence of video clubs on teachers' thinking and practice. *Journal of Mathematics Teacher Education*, *13*(2), 155-176.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221.

Wake, D., Dailey, D., Cotabish, A., & Benson, T. (2017). The Effects of Virtual Coaching on Teacher Candidates' Perceptions and Concerns Regarding On-Demand Corrective Feedback. *Journal of Technology and Teacher Education*, *25*(3), 327–357.

Walker, E., Rummel, N., & Koedinger, K. R. (2011). Using automated dialog analysis to assess peer tutoring and trigger effective support. In *International Conference on Artificial Intelligence in Education* (pp. 385-393). Springer, Berlin, Heidelberg.

Walkoe, J. (2015). Exploring teacher noticing of student algebraic thinking in a video club. *Journal of Mathematics Teacher Education*, *18*, 523–550.

Walkoe, J. & Levin, D. M. (2018). Using technology in representing practice to support preservice teachers' quality questioning: The roles of noticing in improving practice. *Journal of Technology & Teacher Education*, *26*(1), 127–147.

Weston, T. L., Kosko, K. W., Amador, J. M., & Estapa, A. T. (2018). Preservice teachers' questioning: Comparing platforms for practice-based teacher education. *Journal of Technology and Teacher Education*, *26*(1), 149–172.

Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester Jr. (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1053-1098). Greenwich, CT: Information Age Publishing.

Wood, T (1998). Alternative patterns of communication in mathematics classes: Funneling or focusing? In H. Steinbring, M. G. Bartolini Bussi, & A. Sierpinska (Eds.), *Language and Communication in the Mathematics Classroom* (pp. 167-78). Reston, Va.: NCTM.

Xhakaj, F., Aleven, V., & McLaren, B. M. (2017, September). Effects of a teacher dashboard for an intelligent tutoring system on teacher knowledge, lesson planning, lessons and student learning. In *European Conference on Technology Enhanced Learning* (pp. 315-329). Springer, Cham.

Zehner, F., Sälzer, C., & Goldhammer, F. (2016). Automatic coding of short text responses via clustering in educational assessment. *Educational and psychological measurement, 76*(2), 280-303.