

Mikeska, J. N., Steinberg, J., Lottero-Perdue, P. S., Cisterna D., Kinsey, D., & Howell, H. (2023). Using simulated classrooms to examine elementary teachers' perceptions about, attention to, and use of formative feedback to improve their ability to facilitate science discussions. *Contemporary Issues in Technology and Teacher Education*, 23(1), 48-83.

# Using Simulated Classrooms to Examine Elementary Teachers' Perceptions About, Attention to, and Use of Formative Feedback to Improve Their Ability to Facilitate Science Discussions

Jamie N. Mikeska and Jonathan Steinberg  
*ETS*

Pamela S. Lottero-Perdue  
*Towson University*

Dante Cisterna, Devon Kinsey, and Heather Howell  
*ETS*

The purpose of this research study was to examine in-service teachers' perceptions about, attention to, and use of two different types of written formative feedback designed to support them in developing their ability to engage in one core teaching practice: facilitating discussions that engage students in scientific argumentation. This core teaching practice has been nominated as important for building students' scientific literacy and has been perennially difficult for teachers to learn how to engage in successfully. This study used an online, simulated classroom made up of five upper elementary student avatars as the practice-based space where the participating in-service teachers facilitated two science discussions. Following the first discussion, each teacher received one of two types of written formative feedback (namely, specific feedback or scoring level feedback) based on their discussion. Findings indicated that the teachers perceived both types of feedback as useful and strongly attended to areas of improvement suggested in the written feedback, regardless of feedback condition. Study results suggest that the use of specific or scoring level feedback can serve as one valuable mechanism to support improvement when using online simulations with elementary science teachers.

During the last decade, there has been an increasing uptake in the use of practice-based teacher education approaches within both preservice and in-service teacher learning contexts (Benedict et al., 2016; Forzani, 2014; McDonald et al., 2013; Peercy & Troyan, 2017). Practice-based teacher education involves situating teachers' learning in aspects of the work of teaching using various pedagogies of practice to develop teachers' abilities to engage in core teaching practices (Grossman, Compton et al., 2009; Kazemi & Wæge, 2015; Lampert, 2010). One such pedagogy – approximations of practice – provides opportunities for teachers to practice in situations of reduced complexity (Grossman, Hammerness et al., 2009).

In science education, these approximations have included (a) face-to-face peer rehearsals, where teachers try out an instructional strategy with their peers acting as K-12 students (Benedict-Chambers et al., 2020; Davis et al., 2017; Masters, 2020), and (b) online simulated classrooms, where teachers interact with digitally animated student avatars controlled on the backend using a human-in-the-loop who has been trained to respond in real-time as K-12 students (Lee et al., 2018; Levin et al., 2019; Mikeska & Howell, 2020; Straub et al., 2015). Research has suggested that such approximations can develop teachers' beliefs, understandings, and skills (Benedict-Chambers, 2016; Dotger et al., 2010; Mikeska & Howell, 2021). Yet, gaps remain in understanding the mechanisms that support science teachers' learning when using approximations, especially how simulated classrooms can be used productively.

One area ripe for examination is the use of written formative feedback to support teachers in learning from their engagement within approximations. Research has indicated that teachers are more likely to develop their ability to engage in core teaching practices when they are provided with “high-quality opportunities to practice, coupled with support and feedback” (Benedict et al., 2016, p. 2), which supports their reflection on teaching challenges they experience (Grossman, Hammerness et al., 2009). The simulated classroom is a novel tool, where teachers can practice, receive and reflect on feedback, and then use what they learned to refine their teaching skills through repeated practice.

Using feedback to reflect on their instructional practice provides an opportunity for teachers to activate their professional noticing, which is an essential component of science teaching expertise (Chan et al., 2021; Haverly et al., 2020; Luna & Sherin, 2017; Radloff & Guzey, 2017; Tekkumru-Kisa et al., 2018). While research has suggested that timely, actionable, and consistent feedback related to learning outcomes supports improvement (Brookhart, 2008; Hattie & Timperley, 2007; McManus, 2008; Wiggins, 2012), we found no studies that examined feedback conditions within the context of using simulated classrooms as approximations of practice to support science teacher learning.

## **Research Purpose and Questions**

The purpose of this study was to examine in-service elementary teachers' perceptions about, attention to, and use of two different types of written

formative feedback designed to support them in developing their ability to engage in one core teaching practice: facilitating discussions that engage students in scientific argumentation. This core teaching has been nominated as important for building students' scientific literacy (Kloser, 2014; National Research Council, 2007; Windschitl et al., 2012) and ensuring that students meet the recent science standards (NGSS Lead States, 2013).

This teaching practice has also been perennially difficult for teachers to learn how to engage in successfully and, as such, requires focused and scaffolded learning opportunities to build teachers' expertise (Marco-Bujosa et al., 2017; McNeill et al., 2016; Osborne et al., 2013). The participating in-service teachers in this study facilitated these science discussions within an online, simulated classroom, shown in Figure 1, that is made up of five upper elementary student avatars.

**Figure 1**  
*Mursion's Upper Elementary Simulated Classroom*



Image provide by Mursion, Inc.

Findings from this exploratory study can generate hypotheses about the nature of written formative feedback that may be most useful to support teachers' learning within practice-based teacher education, especially when using novel tools like simulated classrooms as the context for their approximations of practice. Teacher educators, professional development facilitators, and educational researchers can use the study results to consider the key characteristics of written formative feedback that are useful to support teachers' learning when they engage in core teaching practices like facilitating argumentation-focused discussions. The study's research questions (RQs) are as follows:

- RQ 1: How do in-service elementary teachers understand and perceive the usefulness of written formative feedback to help them improve their ability to facilitate argumentation-focused discussions? How do these perceptions differ, if at all, between feedback conditions (specific vs. scoring level)? (Perceptions of Feedback Understanding and Usefulness)
- RQ 2: What do in-service elementary teachers report attending to when reviewing their written formative feedback? How do they report using the feedback to inform the next steps in their planning and preparation? How does their attention and use differ, if at all, across the two feedback conditions? (Attention to and Use of Feedback)
- RQ 3: After reviewing formative feedback about their first discussion performance, do in-service elementary teachers improve in their ability to facilitate argumentation-focused discussions? If so, does their improvement relate to the nature of the feedback they received? (Changes in Teachers' Ability to Facilitate Discussions)
- RQ 4: To what extent is there alignment between areas of improved practice and the features of high-quality argumentation-focused discussions that in-service elementary teachers report attending to in their written formative feedback? How does this alignment differ, if at all, across the two feedback conditions? (Alignment between Discussion Improvement and Attention to Feedback)

## **Theoretical Framework**

This study was grounded in two theoretical perspectives. The first is a theory of practice-based teacher education, which focuses on the importance on grounding teachers' learning in the work of teaching. The simulated classroom used in this study served as the practice space whereby study participants engaged in practice-based learning opportunities to build their instructional practice. The second perspective focuses on understanding the mechanisms by which instructional feedback can be used to support teacher learning. This study examined how study participants used and attended to written formative feedback to support their learning of facilitating argumentation-focused discussions.

### **Practice-Based Teacher Education**

Teacher education aims to develop a set of knowledge, competencies, and practices to prepare teachers to teach effectively (Korthagen & Kessels, 1999). While different approaches have been proposed for building teachers' theoretical foundation and their ability to enact classroom practices, practice-based teacher education emphasizes preparing teachers to enact specific instructional practices that occur as part of the day-to-day activities in which teachers and students engage (Borko et al., 2011; Fishman et al., 2017; Forzani, 2014; Pella, 2015). High-quality practices, such as facilitating discussions that engage students in scientific argumentation, are key in teacher education curricula. They can create a milieu with rich and multiple interactions with social and cognitive

experiences that support student thinking and learning (Martin & Dismuke, 2018). Although consensus is growing as to the importance for teachers to learn in and from practice, Zeichner (2010) pointed out that little agreement exists about the conditions for teacher learning that enable learning while practicing instruction.

To address this gap, there have been efforts to develop research-based frameworks that guide the design of practice-based teacher education. Grossman, Hammerness et al. (2009) identified three key pedagogies of practice for learning to enact practices: representations, decompositions, and approximations of practice. Representations of practice help make aspects of specific instructional practices visible to novices and attend to the different ways that practice is represented in professional education. Decompositions of practice consist of breaking down a single practice into its constituent parts for the purposes of teaching and learning.

Approximations of practice refer to opportunities for novices to engage in trying instructional skills that are proximal to the practices of the teaching profession, such as eliciting student ideas in a peer teaching rehearsal or facilitating a discussion with student avatars in an online simulated classroom. Approximations of practice offer teachers opportunities to tryout parts of teaching in safe, reduced complexity, and supportive settings where feedback from peers or teacher educators is key to facilitate teacher learning.

Overall, providing opportunities for teachers to interact with and engage in representations, decompositions, and approximations of practice can support teachers in learning how to enact high-leverage or core teaching practices in classrooms (e.g., Ball & Forzani, 2009; Franke et al., 2006; Kazemi et al., 2007; Kloser, 2014; Sleep et al., 2007). These core teaching practices can be implemented through different curricula, contexts, or instructional approaches and teachers can begin to master them within their initial stages of preparation and continue to do so as they gain expertise and become more experienced practitioners as in-service teachers. Such practices allow teachers to learn more about students and teaching, preserve the integrity and complexity of teaching, and when used well, can improve student achievement.

In this study, we used an online simulated classroom as a practice space to engage in-service elementary teachers in two approximations of practice, where they tried out facilitating a science discussion that engaged five student avatars in argumentation. This practice space provides a way to conduct exploratory research on specific aspects that can support teacher learning – in this case, the type of written formative feedback that teachers receive and use as they build toward improvement in a core teaching practice. Earlier research in this area has investigated the use of this tool to support preservice teacher learning, with more limited work considering such use with an in-service teacher population (Mikeska & Howell, 2021; Mikeska & Lottero-Perdue, 2022; Mikeska et al., 2021).

## **Instructional Feedback to Support Learning**

For this study, feedback is defined as “information provided by an agent (e.g., teacher, peer) regarding aspects of one’s performance or understanding...” (Hattie & Timperley, 2007, p. 81) and is designed to support future improvement. The primary goal of feedback is to help one see the connection between their current understanding or performance and an expected outcome and to illustrate what one can do to move closer to a specified learning goal (Landauer et al., 2009; Wiliam, 2013). As noted by many scholars, feedback is a critical lever in improving learning outcomes (Hattie, 2009; Hattie & Timperley, 2007; Lipnevich & Smith, 2018).

Instructional feedback is seen as an integral part of ongoing formative assessment, where learners are provided with information to help them understand where they are, where they need to go, and how they can improve (Hattie & Timperley, 2007). However, to influence learning outcomes, learners need to engage with and use the feedback to improve (de Kleijn, 2021).

Much research in this area has focused on examining the provision of instructional feedback to support student learning within K-12 classroom settings (Lipnevich & Smith, 2018; Van der Kleij & Lipnevich, 2021). Findings have indicated that (a) there is variability in whether students perceive the feedback they receive as useful (Handley et al., 2011; Jonsson & Panadero, 2018; Rakoczy et al., 2013) and (b) there are numerous factors that relate to whether and how K-12 students engage with instructional feedback, such as the feedback’s content, when and how the feedback is provided, the amount and specificity of the feedback provided, and students’ emotions (Barton et al., 2016; Coogle et al., 2018; Goetz et al., 2018; Harris et al., 2014; Scheeler et al., 2004; Van der Kleij & Lipnevich, 2021).

More recently, research studies have examined how instructional feedback can support preservice and in-service teacher learning within teacher education and professional development settings (Akerson et al., 2017; Aljadef-Abergel et al., 2017; Chizhik & Chizhik, 2018; Ellis & Loughland, 2017; Garet et al., 2017; Kwok, 2018; McGraw & Davis, 2017; McLeod et al., 2019; Prilop et al., 2020). Similar to studies in the student space, teachers’ perceptions of feedback across these varied sources tend to be positive in nature.

For example, findings from one study indicated that most teachers strongly valued feedback from students when feedback was targeted, actionable, and could be implemented immediately (Keiler et al., 2020). In another study, findings indicated that preservice teachers responded positively to video and email feedback they received from coaches and showed an improvement in their ability to engage in recommended instructional practices, such as giving students choices (McLeod et al., 2019).

Studies addressing the use and understanding of instructional feedback to support teacher development have also examined some key factors that

influence when, how, and why teachers attend to and use feedback. One study examined how the timing of feedback relates to preservice teachers' ability to engage in specific teaching skills (Aljadeff-Abergel et al., 2017) and showed that feedback received immediately prior to the next lesson resulted in the greatest impact on their teaching performance.

Another study investigated variation in the mode used to provide the feedback with findings indicating that the preservice teachers valued both modes of feedback delivery, although those who received video-annotated feedback from their coaches increased their growth mindset and outperformed those who received traditional coaching via oral feedback on the standardized performance assessment used (Chizhik & Chizhik, 2018). A third study examined the nature of the feedback provided to preservice teachers by their mentor teachers and found that mentors who taught in schools that focused on reflective practice were more likely to provide feedback that supported the preservice teachers in considering and reflecting on the strategies they were using and engaging in collaborative problem-solving (McGraw & Davis, 2017).

In this study, we leveraged these earlier findings to examine whether there was variability in how study participants perceived and used two different types of feedback as they worked to improve their ability to facilitate argumentation-focused discussions. This study also addressed one of the gaps in the feedback literature – examining how learners' uptake of feedback relates to changes in their actual performance. A few studies in the teacher learning space have started to address this gap (Aljadeff-Abergel et al., 2017; Garet et al., 2017; McLeod et al., 2019), but additional studies are needed to better understand this relationship, especially in terms of key factors that may influence this relationship, such as the nature of the feedback provided.

## **Research Methods**

### **Sample**

A convenience sample of 15 in-service elementary teachers who were teaching in public-school systems located in the Mid-Atlantic area of the United States participated in this study. We sent out emails to local school district leaders and in-service teachers we had worked with on previous projects to recruit study participants. Each interested participant completed a short, online survey to ensure that they met the study requirements for participation.

As shown in Table 1, most participants had an undergraduate degree in elementary education and described themselves as female and White, which is similar to demographics of elementary teachers in the United States (Banilower et al., 2018). On average, they reported teaching science and engineering about 5 to 6 hours per week, respectively, and over the last 3 years they spent, on average, about 8 hours in professional learning experiences in engineering. While some of those experiences may have focused on learning about scientific argumentation, none of the participants described participating in a professional development program focused on this topic. In terms of overall elementary teaching

experience, on average participants had been teaching about 10 years ( $M = 9.8$  years;  $SD = 5.7$  years) with a range of 3 to 22 years. No participants had experience using simulated classrooms.

**Table 1**  
*Characteristics of Participant Sample*

Characteristic	Participant Sample <i>f</i> (%)
<b>Gender</b>	
Female	13 (87%)
Male	2 (13%)
<b>Race/Ethnicity</b>	
White	13 (87%)
Hispanic/Latino	2 (13%)
<b>Undergraduate Degree*</b>	
Early Childhood Education	1 (7%)
Elementary Education	11 (73%)
Special Education	2 (13%)
Other	4 (27%)
<b>Professional Learning Experience Lead Role*</b>	
In-school	3 (20%)
Out-of-school	2 (13%)
State/national	1 (7%)
Characteristic	<b><i>M</i> (<i>SD</i>), Min-Max</b>
<b>Teaching Hours/Week</b>	
Science	5.3 (3.8), 2-15
Engineering	6.5 (5.8), 2-20
<b>Professional Learning Hours/Last 3 Years</b>	
Engineering	8.2 (8.4), 0-26
<i>Note:</i> Categories with 0% participants are not shown here.	
* Respondent could select more than one option.	

### Discussion Performance Task

In this study, each participant had an opportunity to facilitate two discussions (once in fall 2019 and again in winter 2020, approximately 5 months between sessions) in the Mursion® upper elementary simulated classroom using a performance task called Design a Shoreline (abbreviated as “Shoreline”; Lottero-Perdue, Mikeska et al., 2020). In the Shoreline task, teachers are challenged to facilitate a discussion with the student avatars to meet the following student learning goal: “...to work collaboratively to critique and revise each team’s initial ideas about their design performance and improvement based on the outcomes from an engineering design challenge” (p. 3). The engineering design challenge involves having student teams redesign a piece of shoreline to reduce erosion into a bay and create more terrapin habitat while considering cost (Lottero-Perdue, Haines et al., 2020).



The Shoreline task document shared with the study participants provided the teachers with key information they needed to prepare for facilitating the discussion, such as the discussion's student learning goal, any instructional activities the students in the class completed prior to the discussion, background information about the engineering design challenge, copies of the design and design testing images (see Figure 2), and copies of the written work students produced as part of the engineering design challenge. The Shoreline task was created such that each of the three student teams (a) had a design with both strengths and weaknesses and (b) had written statements about design performance and improvement that had positive features and areas that could be improved.

For this study, prior to each data collection period, the simulation specialist engaged in a scaffolded and structured training process to learn about the shoreline design challenge, each team's initial designs and ideas related to the challenge, and when and how the students' ideas could change during the discussion so they could enact these student thinking profiles with consistency across the Shoreline discussions that the in-service teachers facilitated. For a fuller discussion of the Shoreline task, see Lottero-Perdue, Mikeska et al. (2020).

### **Data Collection**

Data collection included a compilation of survey, interview, and observational measures. Similar measures were used in a previous research study with piloting completed to ensure that the questions elicited the expected response from teachers. First, each teacher completed a background survey about their personal demographics, teaching experience, and educational background. Then, they prepared for and facilitated the Shoreline discussion in the simulated classroom for up to 25 minutes, which was video recorded.

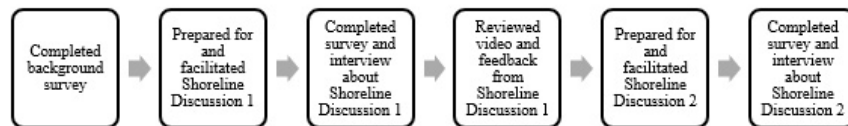
Immediately after their first discussion, they completed an online survey and participated in a semistructured interview about their experience. Within 48 hours, they received a copy of their video-recorded discussion. Approximately 4 months later they received written formative feedback about their first discussion and used it to prepare for and facilitate a second discussion in the simulated classroom about a month later, which was also video recorded, using the same Shoreline task. After their second Shoreline discussion, each teacher completed another online survey and participated in another interview about their second discussion.

Figure 3 illustrates the key steps in the data collection process. Each participant received an honorarium of \$250 for completing the data collection process for the first Shoreline discussion, and \$300 for their participation in the data collection for the second Shoreline discussion, which included their review and use of the formative feedback and video from their first Shoreline discussion to inform their preparation.

**Figure 2**  
*Students' Designs Before and After the Test*



**Figure 3**  
*Steps in the Data Collection Process*



Our research team trained raters, who were practicing or former teachers, on how to write the specific formative feedback to link to each rubric feature and discussion using training protocols and supports from a previous study (GO Discuss Project, 2021; Mikeska & Howell, 2021). The trained raters generated feedback using our four-level scoring rubric; the rubric ranged from Level 1 (beginning practice) to Level 4 (commendable practice).

This rubric was linked to five dimensions – what we referred to as “features” with the teacher participants – of high-quality, argumentation-focused discussions, which we previously identified from our review of the empirical and practitioner literature (Mikeska et al., 2019): (a) attending to student ideas, (b) facilitating a coherent discussion, (c) promoting peer interactions, (d) developing students’ conceptual understanding, and (e) engaging students in argumentation. [Appendix A](#) provides a description of each feature, along with the key scoring indicators for each one.

Each rater engaged in a rigorous training process, including completing multiple learning modules, practice rating example discussions, and meeting with a scoring leader. Each discussion received six scores – one score for each of the five scoring features (scale of 1-4 per feature) and one overall, or aggregate, score that was generated by adding the five feature-level scores (scale of 5-20).

Since this study was focused on examining how the nature of the feedback related to teachers’ feedback perceptions and potential performance improvement, our research team created two different types of written formative feedback for use in this study. Both types of feedback were linked to these five features of high-quality, argumentation-focused discussions. The first type, “specific” feedback, identified specific strengths and areas of need evident in their discussion linked to the five features (Figure 4). The specific feedback identified actual examples from each teacher’s Shoreline Discussion 1 to illustrate strengths and areas of need, as well as provided specific suggestions for how they could improve in each feature.

The second feedback type, “scoring level” feedback (Figure 4), provided the teachers with a numerical rating using our four-level scoring rubric for each of these five features. In addition to the numerical score, for each feature the teacher received bulleted feedback statements stating what teachers at that scoring level tend to do well when facilitating discussions and the typical areas in which they need to improve. While these bulleted statements were not customized to the individual teacher’s discussion, they were customized to the task and scoring rubric to represent patterns we had observed in scoring the set of discussions. After categorizing each participant based on their overall score on the first Shoreline discussion into lower, midlevel, and high scoring groups, we randomly assigned participants from those groups to either the specific ( $n = 7$ ) or scoring level ( $n = 8$ ) condition.

**Figure 4**  
*Example of Feedback Types on Scoring Feature 1*

<b>Specific Feedback</b>
<p style="text-align: center;"><b>Feature 1: Attending to Student Ideas</b></p> <div style="border: 1px solid black; padding: 5px; margin: 5px 0;"> <p><i>Attending to Student Ideas</i> involves being responsive to students and focuses on making sure that the discussion is grounded in students' ideas and that all students are engaged in some meaningful component of the discussion. In providing feedback, we focused on the following three aspects.</p> <ul style="list-style-type: none"> <li>• Incorporating key ideas from the students' work samples into the discussion</li> <li>• Giving every student a chance to have his or her voice heard</li> <li>• Attending to students' ideas shared during the discussion</li> </ul> </div> <p><b>Overall Feedback:</b></p> <p>You incorporated at least one key idea from every group's work samples into the discussion. You probed for meaningful contributions from students throughout the discussion. You elicited at least one meaningful contribution from every student. You made use of students' ideas during the discussion to help move the discussion toward the learning goal.</p> <p><b>Specific Feedback:</b></p> <ul style="list-style-type: none"> <li>• You were able to structure the discussion in a way that allowed all the students' voices to be heard in meaningful ways that went beyond sharing their work samples. Some examples of these meaningful contributions include "Who has some thoughts for them (Carlos and Emily) or feedback for them about their design?" and "I know you like to set goals. Could you maybe set a goal for your design based on Mina's feedback?" Continue to look for opportunities that allow all students to share their ideas in meaningful ways.</li> <li>• While you did a great job giving students opportunities to make their voices heard, consider using the ideas they share to move the conversation forward. For example, at one point during the discussion, Carlos expressed concern that altering his design may "mess up the erosion." Consider what might have happened if you had invited all students to consider his concern during a think-pair-share by asking them to discuss possible solutions that would not negatively impact his design's erosion.</li> </ul>

<b>Feature 1: Attending to Student Ideas</b>							
<p><i>Attending to Student Ideas</i> involves being responsive to students and focuses on making sure that the discussion is grounded in students' ideas and that all students are engaged in some meaningful component of the discussion.</p> <p>Your level is:</p> <table border="1" style="margin: 10px auto; border-collapse: collapse;"> <tr> <td style="padding: 5px; text-align: center;">Level 1 Beginning</td> <td style="padding: 5px; text-align: center;">Level 2 Developing</td> <td style="padding: 5px; text-align: center;">Level 3 Well-Prepared</td> <td style="padding: 5px; text-align: center;">Level 4 Commendable</td> </tr> </table> <p style="text-align: center; margin: 10px 0;">3 out of 4 Well-Prepared</p> <p>Teachers who score at this level typically demonstrate the following <b>strengths</b> while facilitating a discussion.</p> <ul style="list-style-type: none"> <li>• Incorporating all key ideas from every groups' written work into the discussion</li> <li>• Encouraging meaningful contributions from students throughout the discussion</li> <li>• Eliciting at least one meaningful contribution from every student</li> <li>• Making use of students' ideas to move the discussion toward the learning goal</li> </ul> <p>Teachers who score at this level typically focus on the following to <b>improve</b> in attending to student ideas while facilitating a discussion.</p> <ul style="list-style-type: none"> <li>• Encouraging the students to share key ideas from their groups' written work</li> <li>• Encouraging meaningful contributions from students consistently throughout the discussion</li> <li>• Eliciting more than one meaningful contribution from every student</li> <li>• Making use of students' ideas consistently to move the discussion toward the learning goal</li> </ul>				Level 1 Beginning	Level 2 Developing	Level 3 Well-Prepared	Level 4 Commendable
Level 1 Beginning	Level 2 Developing	Level 3 Well-Prepared	Level 4 Commendable				

The online survey used after the second Shoreline discussion asked the teachers to use a 4-point Likert scale (*very easy to understand, somewhat easy to understand, somewhat difficult to understand, and very difficult to understand*) to report on how well they understood the focus and key aspects of each of the five features of high-quality argumentation-focused discussions and how well they understood the formative feedback for each of the five features. Teachers also used a 3-point Likert scale (*very useful, somewhat useful, not useful*) to report on how useful they found the formative feedback for each of the five features of high-quality argumentation-focused discussions. Finally, teachers responded to two open-ended questions within the online survey that focused on what they attended to and how they used the information in the feedback report to inform their next steps:

**Survey Question A:** What did you learn about your past performance from reviewing your written formative feedback report based on your first discussion of the Design a Shoreline task?

**Survey Question B:** What did you decide to do differently in the second discussion after watching your video and reviewing your written formative feedback report from your first discussion of the Design a Shoreline task?

The interview conducted after the second Shoreline discussion aimed to explore teachers' experiences in interpreting and analyzing the formative feedback report as well as how they used the information from the formative report to prepare for the second Shoreline discussion. The interview protocol had teachers describe their main takeaways or learning from the formative feedback and their preparation process for the second Shoreline discussion. Total interview times ranged between 30 and 50 minutes, and all interviews were audio recorded and transcribed for analysis purposes. Interview questions related to participants' perceptions about what they attended to and how they used the formative feedback report are as follows:

**Interview Question A:** What was your main takeaway(s) from the formative feedback report?

**Interview Question B:** In what ways, if at all, did you use the feedback from this performance to inform your planning and preparation for your next task? Please describe any specific aspects of the feedback that you used to inform your planning and preparation.

## **Data Analysis**

Our overall study used a convergent mixed method design (Creswell & Plano Clark, 2011) to analyze teachers' interview and survey responses and their discussion transcripts. We used qualitative analysis approaches to examine in-service teachers' interview and survey responses and characterize their perceptions, attention to, and use of two different types of written formative feedback. We then used quantitative scoring data across the teachers' two Shoreline discussions to determine if and how they improved on the five scoring features. Finally, we compared outcomes from each analysis component to link teachers' perceptions, attention to,

and use of these two feedback formats to their actual scoring performance across time points. This parallel analysis and then direct comparison of qualitative and quantitative data allowed us to discern patterns in alignment between their discussion improvement and attention to feedback features.

### **Analysis for RQ 1**

To answer our first research question about the in-service teachers' understanding and perceptions of the feedback's usefulness, we used descriptive analysis to analyze participants' responses to the task survey after the second Shoreline discussion. For example, we calculated the number and percentage of teachers who selected particular responses on Likert-scale items indicating whether aspects of the feedback reports were *very easy*, *somewhat easy*, *somewhat hard*, or *very hard* to understand and whether the feedback was *very*, *somewhat*, or *not* useful to supporting them in building toward improvement on this teaching practice. We then compared the patterns across the two feedback conditions. We were interested in learning whether the teachers' perceptions about each type of feedback were generally positive in nature and consistent across conditions.

### **Analysis for RQ 2**

To address our second research question, our team used qualitative content analysis (Schreier, 2014) to develop separate coding schemes to characterize the teachers' attention to and use of the feedback. In particular, we analyzed the teachers' responses to the open-ended task survey questions and the interview questions asking them about their main takeaways or learning from the formative feedback report (Survey Question A; Interview Question A) and their use of the formative feedback to inform their planning and preparation for the second discussion (Survey Question B; Interview Question B).

The coding scheme for teachers' attention to the feedback identified the key ideas they stated regarding their main learnings or takeaways from the feedback. However, to characterize teachers' use of the formative feedback, our team developed a three-part coding scheme. The first and broadest code, *Use*, indicated if the participant reported using the feedback. We used two subcodes, *Yes*, the participant used the feedback, or *No*, the participant did not use the feedback. Another code outlined the *Approach* for how teachers used the feedback to prepare for the second discussion (e.g., modified their teaching moves or adjusted the discussion structure). Finally, a *Substance* code identified the nature of the changes they made using the feedback. For this code we identified five subcodes a priori, with each one linked to one of the five rubric features and also identified other ones (e.g., using their discussion time differently) based on what we noticed in the teachers' responses.

For each survey or interview question, two researchers used the relevant coding scheme to code five participant responses together and then independently coded the remaining 10 responses. We then came together to reconcile any areas of disagreement and the independent coding for the

10 responses was used to calculate rater agreement and interrater reliability using the intraclass correlation coefficient (ICC) for each survey or interview question. Coding for each question (Survey Question A, Interview Question A, Survey Question B, and Interview Question B) achieved adequate rater agreement (95%, 85%, 91%, and 90%, respectively) and interrater reliability (ICC = .90, .74, .87, and .88, respectively).

Our team drew upon coding across Survey Question A and Interview Question A to address what the participants attended to in their feedback and coding across Survey Question B and Interview Question B to address how they used their feedback to inform their next steps for planning and preparation of their second Shoreline discussion. For each part, our team identified key patterns using the frequencies by which these codes were applied across teacher participants. Similar to the study's first research question, we compared the patterns in the frequency of codes applied across feedback conditions.

### ***Analysis for RQ 3***

To address the study's third research question, we performed two comparative analyses: (a) examining average overall (aggregate) scores across features for the combined sample ( $n = 15$ ), and (b) examining median scores across features between the specific ( $n = 7$ ) and scoring level ( $n = 8$ ) feedback groups. We chose to use nonparametric approaches, specifically the Wilcoxon Signed-Rank Test (Wilcoxon, 1945) for the first comparative analysis and the median test (Moses, 1952) for the second comparative analysis, since we did not anticipate that we could assume normality of the score distributions due to the study's small sample of convenience. Two-tailed significance at a .05 level was evaluated for each between-subjects analysis.

### ***Analysis for RQ 4***

To answer the study's fourth research question, we used the teachers' task survey and interview responses to compare the features of high-quality argumentation-focused discussion that they reported attending to in their feedback report to their actual areas of improved practice in those features (i.e., improvement in rater scores from their first discussion to their second discussion). For this analysis, we compared the features that participants mentioned attending to in their feedback, as evidenced in their responses to the task survey and interview questions, to the change in scores across features for the sample overall ( $n = 15$ ) and for each of the specific ( $n = 7$ ) and scoring level ( $n = 8$ ) feedback groups.

In this sample, teachers' scores on features from the first Shoreline discussion to the second Shoreline discussion either decreased by 1 score point (change of -1), stayed the same (change of 0), or increased by 1, 2, or 3 points (change of +1, +2, +3, respectively). This comparison of reported attention to one of the five features of high-quality argumentation-focused discussions and change in feature score was evaluated to determine alignment. The two instances of alignment between these aspects occurred when (a) the teacher reported attending to a feature *and* their score for

that feature increased, or (b) the teacher did not report attending to a feature *and* their score for that feature stayed the same or decreased. The two instances of misalignment occurred when (a) the teacher reported attending to a feature *and* their score for that feature stayed the same or decreased, or (b) the teacher did not report attending to a feature *and* their score for that feature increased.

## Findings

### RQ 1: Perceptions

Most participants responded that the feedback they received was *very* ( $n = 12$ ) or *somewhat* ( $n = 2$ ) useful in helping to inform their planning and preparation for the second Shoreline discussion. All participants indicated that the overall descriptions of each scoring feature were *very* or *somewhat* easy to understand (means ranged from 3.60 to 3.93 per feature on a 4-point scale with a score of 4 indicating *very easy to understand*). Regarding the actual feedback about their discussion performance, the response was also positive in that all but one participant (for Features 1 to 4) or all but two participants (for Feature 5) noted that the feedback was *very* or *somewhat* easy to understand (means ranged from 3.67 to 3.80 per feature). Similarly, all participants, except for one on Feature 2, noted that the feedback was *very* or *somewhat* useful in helping them build toward improvement. No differences in patterns were noted across the two feedback groups.

### RQ 2: Attention

Frequencies across all main takeaway and learning codes for all participants, as well as each feedback group, are presented in [Appendix B](#). Findings showed that most teachers (93%) attended to one or more areas of improvement mentioned in their feedback report. Typically, when a teacher reported an area of improvement, they also used language that reflected the descriptions of one of the five scoring rubric features.

For instance, in their task survey question response, Participant 003 (Specific Feedback) noticed that the feedback provided guidance for them to focus on "...the students and how they should be...justifying with evidence and drawing conclusions." This response was coded as both "Areas for Improvement" and "Engaging Students in Argumentation," as the participant identified an area for improvement linked to this feature of the scoring rubric.

The rubric feature most commonly reported by teachers as an important area of improvement was "Developing Students' Conceptual Understanding" (60%). For example, one teacher noted in their task survey question response that the feedback allowed them to think about questions such as "Did you truly develop conceptual understanding? Did you use precise and accurate language?" (Participant 002, Scoring Level Feedback). Other rubric features attended to by teachers included "Promoting Peer Interactions" (47%) and "Engaging Students in Argumentation" (47%), followed by "Facilitating a Coherent Discussion" (40%) and "Attending to Student Ideas" (13%).



One third of participants (33%) also called attention to the areas of strength from their first discussion that were identified in the formative feedback report. For example, on the task survey one teacher noted their areas of strength in Features 3, 4, and 5: “I did well with allowing them to participate in arguments. I was strong with making goals and allowing for discussion for students to understand from their fellow classmates their strengths and weaknesses and ways they can improve their design” (Participant 025, Specific Feedback).

Some responses, however, were general and not linked to a specific rubric feature, such as the response Participant 025 (Specific Feedback) provided in their interview: “I realized that I’m doing pretty well with argumentative discussions.” Other less commonly reported takeaways were linked to how the teacher could have or did use the discussion time differently (20%) or how they learned something related to the rubric features (27%) without identifying specifically what was learned. For example, Participant 013 (Scoring Level Feedback) noted how they “learned that there was opportunity for improvement in all 5 features,” without mentioning anything specific.

As is evident in [Appendix B](#), patterns in attention to the main takeaways and key learnings from the feedback report were consistent across the two feedback groups with the only notable differences in two categories (here, we define “notable” as there being greater than 25 percentage points of difference when comparing the groups). Participants in the scoring level group were more likely to attend to improvements they needed to make in terms of “Facilitating a Coherent Discussion” (Feature 2; 75% of scoring level feedback participants versus no specific feedback participants) and were also more likely to note generically that their learning was linked to the rubric features shown in the feedback report.

In terms of the teachers’ use of the feedback, all teachers (100%) reported using the formative feedback in the planning and preparation for their second Shoreline discussion. [Appendix C](#) presents a summary of the teachers’ reported use of feedback to prepare for the second discussion (Approach code) and the nature of the changes they made using the feedback (Substance code).

With regard to their approach, most teachers (87%) indicated that they made modifications to the teaching moves or activities used in their second Shoreline discussion. For example, Participant 002 (Scoring Level Feedback) identified multiple teaching moves or activities they modified: “encouraging students to evaluate the accuracy and validity of most of their ideas, making progress towards resolving critical student misunderstandings, and using precise and accurate language.” This response also was coded as “Developing Students’ Conceptual Understanding” for the Substance code. Teachers identified approaches linked to adjusting or modifying the notes/outline for the second discussion (53%), the structure of the second discussion (40%), and the goals for the second discussion (20%).

In terms of the substance regarding how the teachers used the feedback, most participants used language that suggested links to one or more of the five scoring rubric features: “Promoting Peer Interactions” (73%),

“Developing Students’ Conceptual Understanding” (73%), “Attending to Student Ideas” (67%), “Facilitating a Coherent Discussion” (67%), and “Engaging Students in Argumentation” (53%). For example, one participant explained how they used the feedback to improve in the “Promoting Peer Interactions” feature (Substance), which also included modifying their teaching moves (Approach):

How am I going to keep [the students] calling on each other and talking the whole time? Okay, the first thing that I say when I set up the norms for the class discussion, “You guys have to keep talking to each other the whole time. That’s what this is going to sound like, just talk to each other. You are speaking to each other about this...” (Participant 009, Scoring Level Feedback)

While most of the patterns noted about participants’ use of the feedback were consistent across feedback groups, there were a few areas where the frequency differed by more than 25 percentage points. First, only participants who received scoring level feedback considered using the feedback to adjust or modify the discussion goal that they planned to address in their second Shoreline discussion. Second, more participants in the specific feedback group noted how they planned to be better in “Attending to Students’ Ideas” (Feature 1) and “Promoting Peer Interactions” (Feature 3) in their preparation for the second Shoreline discussion, while more participants in the scoring level feedback group noted how they planned to use the feedback to better plan for “Developing a Coherent Discussion” (Feature 2).

### RQ 3: Changes

Table 2 shows participants’ average scores within and across the five rubric features at both time points (the first and second Shoreline discussions) across the overall participant sample and by feedback group. Findings indicated that, both within and across features for the total sample, average participant scores improved from Discussion 1 to Discussion 2. At both time points, average scores for Features 4 and 5 were consistently lower than those for Features 1, 2, or 3. Nonetheless, based on the overall score across features, the Wilcoxon Signed-Rank Test detected a significant improvement in average overall scores (Discussion 1:  $M = 13.73$ ,  $SD = 2.60$ ; Discussion 2:  $M = 16.07$ ,  $SD = 2.66$ ;  $p = .01$ ). Additional analyses by feature revealed significant improvement in Feature 1 (Discussion 1:  $M = 2.93$ ,  $SD = 0.59$ ; Discussion 2:  $M = 3.60$ ,  $SD = 0.63$ ;  $p = .01$ ) and Feature 4 (Discussion 1:  $M = 2.47$ ,  $SD = 0.74$ ; Discussion 2:  $M = 3.00$ ,  $SD = 0.76$ ;  $p = .03$ ), but not for Features 2, 3, or 5.

Findings also showed that, both within and across features, average participant scores on the second Shoreline discussion for the specific feedback group ( $n = 7$ ) tended to be higher than those for the scoring level feedback group ( $n = 8$ ). However, based on the overall score across features, the median test did not detect a significant difference in average overall scores between the two feedback conditions at Discussion 1 prior to assignment (Specific Feedback:  $M = 14.14$ ;  $SD = 3.24$ ; Scoring Level Feedback:  $M = 13.38$ ;  $SD = 2.07$ ;  $p = .78$ ) or at Discussion 2 (Specific Feedback:  $M = 17.14$ ;  $SD = 1.77$ ; Scoring Level Feedback:  $M = 15.13$ ;  $SD = 3.04$ ;  $p = .46$ ).

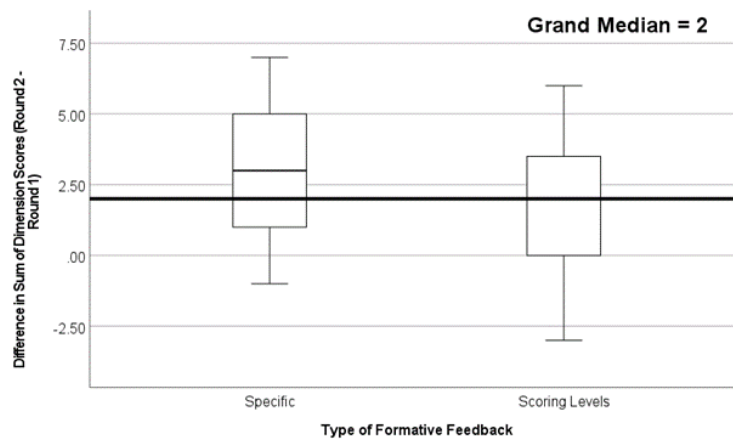
**Table 2**  
Means and Standard Deviations for Participants' Discussion Scores

Feature	Shoreline Discussion 1			Shoreline Discussion 2		
	All Participants (n = 15)	SLF Group (n = 8)	SF Group (n = 7)	All Participants (n = 15)	SLF Group (n = 8)	SF Group (n = 7)
1: Attending to Student Ideas	2.93 (0.59)	2.75 (0.46)	3.14 (0.69)	3.60 (0.63)	3.38 (0.74)	3.86 (0.38)
2: Developing a Coherent Discussion	2.87 (0.64)	2.63 (0.74)	3.14 (0.38)	3.20 (0.68)	3.00 (0.76)	3.43 (0.53)
3: Promoting Peer Interactions	2.87 (0.74)	3.00 (0.53)	2.71 (0.95)	3.33 (0.82)	3.13 (0.83)	3.57 (0.79)
4: Developing Conceptual Understanding	2.47 (0.74)	2.38 (0.52)	2.57 (0.98)	3.00 (0.76)	2.75 (0.89)	3.29 (0.49)
5: Engaging Students in Argumentation	2.60 (0.74)	2.63 (0.74)	2.57 (0.79)	2.93 (0.80)	2.88 (0.64)	3.00 (1.00)
Overall Score	13.73 (2.60)	13.38 (2.07)	14.14 (3.24)	16.07 (2.66)	15.13 (3.04)	17.14 (1.77)

Note. Each feature was scored on a four-level scoring rubric from Level 1 (beginning practice) to Level 4 (commendable practice). The overall score ranged from a low of 5 points (received Level 1 on all five features) to a high of 20 points (received Level 4 on all five features). SLF = Scoring Level Feedback; SF = Specific Feedback.

An additional test for the amount of change in overall scores between feedback groups from Discussion 1 to Discussion 2 also did not produce a significant difference (Specific Feedback:  $M = 3.00$ ;  $SD = 2.89$ ; Scoring Level Feedback:  $M = 2.92$ ;  $SD = 1.75$ ;  $p = .32$ ; see Figure 5). Additional analyses by feature revealed similar results ( $ps > .18$ ). The test could only be performed for Features 2, 4, and 5 due to the fact that all observed values were less than or equal to the median. These results indicate that both feedback types were equally likely to be useful and produce similar degrees of change over time.

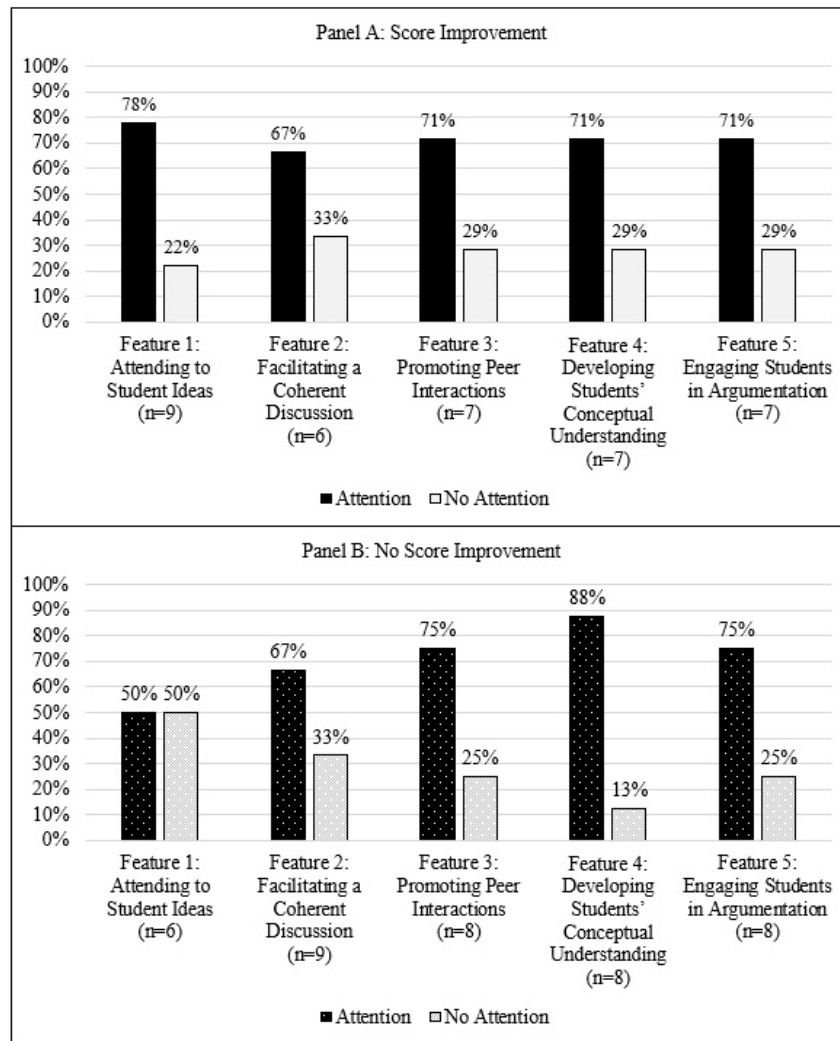
**Figure 5**  
Boxplot of Changes in Overall Scores by Feedback Type from Discussion 1 to Discussion 2



### RQ 4: Alignment

To determine the alignment between actual score improvement in each rubric feature and reported attention to these features, we compared the feature-level codes distributed across the four items assessed for the second research question (Survey Question A, Interview Question A, Survey Question B, and Interview Question B) to the teacher’s discussion score change in each of the five features from their first discussion to their second discussion. As shown across the two panels in Figure 6, of the 15 teachers, 10 (67%), 10 (67%), 11 (73%), 12 (80%), and 11 (73%) showed evidence of attention to Features 1, 2, 3, 4, and 5, respectively, suggesting strong attention to these features when using the feedback.

**Figure 6**  
*Participants’ Attention to Specific Rubric Features With and Without Score Improvement*



Panel A of Figure 6 presents a graph for the subset of teachers who showed score improvement only (per feature) and compares those teachers who reported attention to the feature (alignment) to those who did not report attention to the feature (misalignment). Those who improved their score in Features 1, 2, 3, 4, and 5 reported attending to that feature in 78%, 67%, 71%, 71%, and 71% of teachers' responses, respectively. Taken together, these results suggest that improvement in a specific feature was usually coupled with evidence of attention to that feature in the participant's survey and/or interview responses.

Similarly, Panel B of Figure 6 includes only those teachers who did not improve in their scores for particular features and compares across them to indicate the teachers with no score improvement who did show evidence of attention to this feature in their feedback (misalignment) or did not show evidence of attention to this feature in their feedback (alignment). In this case, the alignment for those with a score that stayed the same or decreased in Features 1, 2, 3, 4, and 5 and who did not report attending to that feature included 50%, 33%, 25%, 13%, and 25% of teachers' responses, respectively.

These findings show that those who did not improve were still more likely to report attention to that feature. As such, these results suggest that most teachers paid attention to these rubric features when preparing for their second Shoreline discussion, although merely attending to a feature did not always translate into actual improvement.

We also examined if there were differences between feedback groups for both those who improved their score in a feature and those who did not improve their score. Overall, we found 36 instances of a teacher improving their score in a feature from their first discussion to their second discussion (e.g., nine teachers improved in Feature 1, six teachers improved in Feature 2, and seven teachers improved in each of Features 3, 4, and 5). Of those 36 instances, 18 occurred in each feedback group. Fourteen of the 18 (78%) separate instances in the scoring level feedback group occurred when a teacher showed evidence of attention to the feature they improved in. The other four instances (22%) occurred when a teacher showed improvement in their score in a feature for their second Shoreline discussion, but there was no evidence in their survey or interview question responses of attending to that feature. Similarly, 12 of the 18 (67%) instances in the specific feedback group occurred when a teacher showed evidence of attention to the feature they improved in while no evidence of attention was shown for the other six instances (33%).

These findings, summarized in Table 3, suggest that a similar pattern occurred in both feedback groups. That is, when score improvement occurred, it was frequently coupled with evidence of attention to that feature in their feedback.

**Table 3**  
*Group Differences for Participants Whose Scores Improved*

Group	Scoring Level Feedback Group ( <i>n</i> = 18 instances)	Specific Feedback Group ( <i>n</i> = 18 instances)
Evidence of Attention to Specific Rubric Features	78%	67%
No Evidence of Attention to Specific Rubric Features	22%	33%
<i>Note.</i> In total, there were 36 instances of a teacher improving their score in a rubric feature (e.g., increasing 1, 2, or 3 points) from their first discussion to their second discussion, with 18 instances per feedback condition.		

Similarly, there was a pattern with the 39 instances in which a teacher did not improve their score in a feature from their first to second discussion (e.g., six teachers did not improve in Feature 1, nine teachers did not improve in Feature 2, and eight teachers did not improve in Features 3, 4, and 5). Of the 39 total instances, 22 occurred in the scoring level feedback group, while 17 occurred in the specific feedback group. When examining teachers who did not improve on a feature, the differences between feedback groups in those who showed evidence of attention to a feature was slightly larger than when examining only those who did improve in the feature. For instance, 14 of the 22 (64%) separate instances in the scoring level feedback group occurred when a teacher showed evidence of attention to the feature they did not improve in. The other eight instances (36%) occurred when a teacher did not show improvement in their score in a feature for their second Shoreline discussion and had no evidence in their survey or interview responses of attending to that feature.

On the other hand, 14 of the 17 (82%) instances in the specific feedback group occurred when a teacher showed evidence of attention to the feature they did not improve in, while no evidence of attention was shown for the other three instances (18%). Again, the overall pattern is similar across feedback groups within this subset. As summarized in Table 4, most cases of no score improvement were coupled with evidence of attention to that feature, suggesting that attention alone is insufficient to making improvement on a particular feature.

## Discussion

Study findings point to the use of written formative feedback – either specific or scoring level – as one productive avenue that can help scaffold teachers’ learning as they engage in approximations of practice using simulated classroom environments. Results suggest that teachers overwhelmingly perceived the written feedback they received as clear and useful, regardless of feedback condition.

**Table 4**  
*Group Differences for Participants Whose Scores Did Not Improve*

<b>Group</b>	<b>Scoring Level Feedback Group (<i>n</i> = 22 instances)</b>	<b>Specific Feedback Group (<i>n</i> = 17 instances)</b>
Evidence of Attention to Specific Rubric Features	64%	82%
No Evidence of Attention to Specific Rubric Features	36%	18%
<i>Note.</i> In total, there were 39 instances of a teacher not improving their score in a rubric feature (e.g., score stays the same or decreases) from their first discussion to their second discussion, with 22 instances in the scoring level feedback group and 17 instances in the specific feedback group.		

Our coding also indicated that their responses consistently reflected the language of the five rubric features, providing additional evidence that they were making sense of the feedback in the ways intended. Findings indicated that, regardless of feedback condition, teachers more often attended to areas of improvement than to areas of strength, and the improvement areas nearly always mapped directly to one or more of the rubric features, providing further evidence that the engagement with the written feedback was as intended.

Results also suggested improvement in the overall sample between the first and second Shoreline discussions. Overall scores and scores for Features 1 and 4 improved significantly between time points. While the improvement in Features 2, 3, and 5 was not statistically significant, it was directionally consistent in the positive direction. The sample size was possibly too small to detect a significant effect. In addition, the groups did not trend differently from one another, suggesting that while more studies may be needed, there is no evidence that the feedback condition assigned was a key differentiator in the level of improvement observed.

Results from the fourth research question suggest more study is needed to understand the path between attention and understanding of feedback and its fruitful application. Findings showed that teachers paid attention to and understood the written feedback and that, on average, the teachers' ability to facilitate argumentation-focused discussions improved. While we had expected to see a clear relationship between the attention teachers gave the feedback and their improvement, we observed a complex relationship: Teachers who did improve showed attention to the improved feature, but teachers who did not improve also showed attention.

We propose several reasonable hypotheses for this pattern. The high levels of attention to the feedback across the board possibly made it impossible to disentangle the relationship between attention and improvement. Certainly, our sample of voluntary participants may have, through self-selection, generated a population that was highly likely to engage with the

feedback. In addition, our scoring measure may have been too coarse to pick up on the level of improvement that might reasonably be expected across only two time points, as the use of whole-number rubric scores precluded us from being able to see improvements within score bands. Similarly, our survey and interview responses provided us only with evidence of attention or lack of evidence. We might not have picked up attention that occurred and influenced planning for the second discussion but that the participant did not share during the survey or interview. Finally, it is plausible that attention to feedback is helpful toward, but insufficient to guarantee, improvement in this teaching skill. Despite our participants' widespread attention to the feedback, feedback alone may not offer sufficient scaffolding for improvement.

Our findings on teachers' positive perceptions about the usefulness of feedback that helps them understand their strengths and areas of improvement and potential next steps they can take align with previous research. As noted in other research, it is important that learners perceive the feedback they receive – no matter the source, nature, or timing of the feedback – as useful and relevant. If not, then they will be unlikely to engage with the feedback and use it to inform their next steps (Jonsson & Panadero, 2018; Van der Kleij & Lipnevich, 2021).

In this study, all the teachers reported that they found the feedback to be useful. Further, they were able to identify certain feedback characteristics, particularly the identification of strengths and specific improvement areas, as being useful in helping them prepare and build toward improvement in this teaching practice. This finding underscores the point that feedback needs to be used by the learner to enact change. It also highlights how certain feedback aspects are particularly useful, such as identifying key areas for improvement (Hattie & Timperley, 2007; Van der Kleij et al., 2015), which was a characteristic included in both types of feedback used in this study.

Findings also aligned with previous research suggesting that feedback can be an important mechanism to support positive change in teachers' ability to engage in specific instructional practices. One of the primary goals of feedback is directly to help learners improve their performance on a specific task or improve their learning. As noted earlier, only a few studies have started to examine the ways learners' uptake of feedback relates to changes in their actual performance (Aljadef-Abergel et al., 2017; Garet et al., 2017; McLeod et al., 2019). This study directly adds to this aspect of the literature, as our findings show that the teachers used the written feedback – particularly the improvement areas identified within the feedback – to focus their preparation and, on average, the teachers showed significant gains in their overall rubric scores during their second Shoreline discussion.

Finally, our findings suggest few meaningful differences between the feedback condition groups. Interpretations of these results require caution due to the small samples assigned to each condition, but the finding is somewhat counter-intuitive given prior research suggesting that feedback is generally more effective when it is more specific. The finding is also potentially meaningful for the field, as it suggests that effective feedback may not need to be linked directly to an individual's performance to be



meaningful and useful, suggesting more efficient and cost-effective methods for generating such feedback may be possible.

We note here that the nature of the scoring level feedback condition may be critical. The scoring level feedback, while not linked specifically to the teacher's own discussion, reflected a research-based accumulation of highly specific observations and actionable suggestions based on the five scoring rubric features. It may be that because the scoring level statements were tied so closely to these features, they were sufficiently specific and actionable for teachers to use them productively even without ties to their specific discussions. This suggests an additional affordance of teacher training in controlled environments such as simulations, as using a common task and rubric features controls variation in ways that make such feedback possible.

### **Limitations**

This study has a few key limitations. First, the exploratory nature of the study and the study's small sample of convenience not only limits any claims of generalizability across the elementary in-service teacher population nationwide, but also limits our ability to make strong claims about feature level improvement and differences between feedback conditions. However, the findings can still point to working hypotheses that can be examined on a larger scale to understand better the ways in which specific feedback characteristics are used to support teacher learning. Second, this study did not collect data to examine how these teachers used the feedback to engage in facilitating argumentation-focused science discussions in actual classrooms. The teachers' ability to facilitate such discussions in actual classrooms may be different. Third, this study did not include a control group who did not receive any feedback to better understand the role of written feedback in supporting teacher improvement, a gap which should be addressed in subsequent studies. Finally, in this study we did not investigate other factors, such as teachers' understanding of engineering or their experience receiving and responding to feedback, which may have impacted teachers' use of the written feedback.

### **Conclusion**

Findings from this study point to the importance of written feedback as a viable mechanism to support teacher learning. Most importantly, study results suggest that both specific and scoring level feedback provide potential positive benefits and can serve as a productive mechanism in helping teachers identify key areas of improvement and provide them with ideas for strengthening specific aspects of their practice. These findings are especially promising considering the ambitious nature of this core teaching practice and how previous research has indicated that this practice is one that is difficult for teachers to learn how to do well. Future research can examine if and how teachers can apply their improvement in the simulated classroom to their instruction in actual elementary classrooms and explore other factors that support teacher learning, such as the feedback timing and the use of coaching.

Additional research also can explore how to develop automated scoring and feedback models to scale the provision of instructional feedback when used in combination with simulated classrooms in practice-based teacher education. Finally, to extend the work from this study, researchers should explore the use of such feedback conditions with a larger, more representative sample of in-service teachers and should expand exploration to include teachers across the professional continuum, especially preservice teachers in teacher education programs, to better understand the specific affordances and potential challenges of using these feedback features to support teacher learning.

### Author Note

This study was funded by Towson University's School of Emerging Technologies and ETS's internal allocation funding. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s).

### References

- Akerson, V. L., Pongsanon, K., Park Rogers, M. A., Carter, I., & Galindo, E. (2017). Exploring the use of lesson study to develop elementary preservice teachers' pedagogical content knowledge for teaching nature of science. *International Journal of Science and Mathematics Education, 15*(2), 293-312. <https://doi.org/10.1007/s10763-015-9690-x>
- Aljadef-Abergel, E., Peterson, S. M., Wiskirchen, R. R., Hagen, K. K., & Cole, M. L. (2017). Evaluating the temporal location of feedback: Providing feedback following performance vs. prior to performance. *Journal of Organizational Behavior Management, 37*(2), 171-195. <https://doi.org/10.1080/01608061.2017.1309332>
- 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. <https://doi.org/10.1177/0022487109348479>
- Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+*. Horizon Research, Inc.
- Barton, E. E., Fuller, E. A., & Schnitz, A. (2016). The use of email to coach preservice early childhood teachers. *Topics in Early Childhood Special Education, 36*(2), 78-90. <https://doi.org/10.1177/0271121415612728>
- Benedict, A., Holdheide, L., Brownell, M., & Foley, A. M. (2016). *Learning to teach: Practice-based preparation in teacher education*. American Institutes for Research.
- Benedict-Chambers, A. (2016). Using tools to promote novice teacher noticing of science teaching practices in post-rehearsal discussions. *Teaching and Teacher Education, 59*, 28-44. <https://doi.org/10.1016/j.tate.2016.05.009>

Benedict-Chambers, A., Fick, S. J., & Arias, A. M. (2020). Preservice teachers' noticing of instances for revision during rehearsals: A comparison across three university contexts. *Journal of Science Teacher Education*, 31(4), 435-459. <https://doi.org/10.1080/1046560X.2020.1715554>

Borko, H., Koellner, K., Jacobs, J., & Seago, N. (2011). Using video representations of teaching in practice-based professional development programs. *ZDM*, 43(1), 175-187. <https://doi.org/10.1007/s11858-010-0302-5>

Brookhart, S. M. (2008). *How to give effective feedback to your students*. ASCD Press.

Chan, K. K. H., Xu, L., Cooper, R., Berry, A., & van Driel, J. H. (2021). Teacher noticing in science education: Do you see what I see? *Studies in Science Education*, 57(1), 1-44. <https://doi.org/10.1080/03057267.2020.1755803>

Chizhik, E., & Chizhik, A. (2018). Value of annotated video-recorded lessons as feedback to teacher-candidates. *Journal of Technology and Teacher Education*, 26(4), 527-552. <http://www.learntechlib.org/p/182175/>

Coogle, C. G., Ottley, J. R., Storie, S., Rahn, N. L., & Kurowski-Burt, A. (2018). Performance-based feedback to enhance preservice teachers' practice and preschool children's expressive communication. *Journal of Teacher Education*, 1-15. <https://doi.org/10.1177/0022487118803583>

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage.

Davis, E. A., Kloser, M., Wells, A., Windschitl, M., Carlson, J., & Marino, J. (2017). Teaching the practice of leading sense-making discussion in science: Science teacher educators using rehearsals. *Journal of Science Teacher Education*, 28(3), 275-293. <https://doi.org/10.1080/1046560X.2017.1302729>

de Kleijn, R. A. (2021). Supporting student and teacher feedback literacy: an instructional model for student feedback processes. *Assessment & Evaluation in Higher Education*, 1-15. <https://doi.org/10.1080/02602938.2021.1967283>

Dotger, S., Dotger, B. H., & Tillotson, J. (2010). Examining how preservice science teachers navigate simulated parent-teacher conversations on evolution and intelligent design. *Science Education*, 94(3), 552-570. <https://doi.org/10.1002/sce.20375>

Ellis, N. J., & Loughland, T. (2017). 'Where to next?' Examining feedback received by teacher education students. *Issues in Educational Research*, 27(1), 51-63.

Fishman, E. J., Borko, H., Osborne, J., Gomez, F., Rafanelli, S., Reigh, E., Tseng, A., Million, S., & Berson, E. (2017). A practice-based professional development program to support scientific argumentation from evidence in the elementary classroom. *Journal of Science Teacher Education*, 28(3), 222-249. <https://doi.org/10.1080/1046560X.2017.1302727>

Forzani, F. M. (2014). Understanding “core practices” and “practice-based” teacher education: Learning from the past. *Journal of Teacher Education*, 65(4), 357-368. <https://doi.org/10.1177/0022487114533800>

Franke, M., Grossman, P., Hatch, T., Richert, A. E., & Schultz, K. (2006, April 7-11). *Using representations of practice in teacher education*. [Paper presentation.] American Educational Research Association Annual Meeting, San Francisco, CA.

Garet, M. S., Wayne, A. J., Brown, S., Rickles, J., Song, M., & Manzeske, D. (2017). *The impact of providing performance feedback to teachers and principals* (NCEE 2018-4001). National Center for Education Evaluation and Regional Assistance.

GO Discuss Project (2021). *Scoring*. Qualitative data repository. <https://doi.org/10.5064/F6NJU10I>

Goetz, T., Lipnevich, A., Krannich, M., & Gogol, K. (2018). Performance feedback and emotions. In A. Lipnevich & J. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 554–574). Cambridge University Press. <https://doi.org/10.1017/9781316832134.027>

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. <https://doi.org/10.1177/016146810911100905>

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice*, 15(2), 273-289. <https://doi.org/10.1080/13540600902875340>

Handley, K., Price, M., & Millar, J. (2011). Beyond “doing time”: Investigating the concept of student engagement with feedback. *Oxford Review of Education*, 37, 543–560. <https://doi.org/10.1080/03054985.2011.604951>

Harris, L. R., Brown, G. T. L., & Harnett, J. A. (2014). Understanding classroom feedback practices: a study of New Zealand student experiences, perceptions, and emotional responses. *Educational Assessment, Evaluation and Accountability*, 26, 107–133. <https://doi.org/10.1007/s11092-013-9187-5>

Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P.M. Johnston, & M. Rees (Eds.), *Tertiary Assessment & Higher Education*

*Student Outcomes: Policy, Practice & Research* (pp. 259-275). Ako Aotearoa.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>

Haverly, C., Calabrese Barton, A., Schwarz, C. V., & Braaten, M. (2020). “Making space”: How novice teachers create opportunities for equitable sense-making in elementary science. *Journal of Teacher Education*, 71(3), 63-79. <https://doi.org/10.1177/0022487118800706>

Jonsson, A., & Panadero, E. (2018). Facilitating students’ active engagement with feedback. In A. Lipnevich & J. Smith (Eds.), *The Cambridge handbook of instructional feedback* (Cambridge handbooks in psychology; pp. 531–553). Cambridge University Press. <https://doi.org/10.1017/9781316832134.026>

Kazemi, E., & Wæge, K. (2015). Learning to teach within practice-based methods courses. *Mathematics Teacher Education and Development*, 17(2), 125–145.

Kazemi, E., Lampert, M., & Ghouseini, H. (2007). *Conceptualizing and using routines of practice in mathematics teaching to advance professional education* (Report). Spencer Foundation.

Keiler, L. S., Diotti, R., Hudon, K., & Ransom, J. C. (2020). The role of feedback in teacher mentoring: How coaches, peers, and students affect teacher change. *Mentoring & Tutoring: Partnership in Learning*, 28(2), 126-155. <https://doi.org/10.1080/13611267.2020.1749345>

Korthagen, F. A., & Kessels, J. P. (1999). Linking theory and practice: Changing the pedagogy of teacher education. *Educational Researcher*, 28(4), 4-17. <https://doi.org/10.3102/0013189X028004004>

Kloser, M. (2014). Identifying a core set of science teaching practices: A delphi expert panel approach. *Journal of Research in Science Teaching*, 51(9), 1185-1218. <https://doi.org/10.1002/tea.21171>

Kwok, A. (2018). Promoting “quality” feedback: First-year teachers’ self-reports on their development as classroom managers. *Journal of Classroom Interaction*, 53(1), 22–36. <http://www.jstor.org/stable/45373100>

Lampert, M. (2010). Learning teaching in, from, and for practice: What do we mean? *Journal of Teacher Education*, 61(1-2), 21-34. <https://doi.org/10.1177/0022487109347321>

Landauer, T. K., Lochbaum, K. E., & Dooley, S. (2009). A new formative assessment technology for reading and writing. *Theory into Practice*, 48(1), 44-52. <https://doi.org/10.1080/00405840802577593>

Lee, C. W., Lee, T. D., Castles, R., Dickerson, D., Fales, H., & Wilson, C. M. (2018). Implementation of immersive classroom simulation activities in a mathematics methods course and a life and environmental science course. *Journal of Interdisciplinary Teacher Leadership*, 2(1), 3-18. <https://par.nsf.gov/biblio/10105492>

Levin, D., Grosser-Clarkson, D., Galvez Molina, N., Haque, A., Fleming, E., & Chumbley, A. (2019, February 19-21). *Pre-service middle school science teachers' practices of leading discussion with virtual avatars*. [Paper presentation.] Simulations in Teacher Education Conference, Louisville, KY.

Lipnevich, A. A., & Smith, J. K. (Eds.). (2018). *The Cambridge handbook of instructional feedback*. Cambridge University Press.

Lottero-Perdue, P. S., Haines, S., Baranowski, A. & Kenny, P. (2020). Designing a model shoreline: Creating habitat for terrapins and reducing erosion into the bay. *Science and Children*, 57(7), 40-45. <https://www.jstor.org/stable/27045241>

Lottero-Perdue, P. S., Mikeska, J., & Orlandi, E. (2020, June 22-26). *Development and teacher perceptions of an avatar-based performance task for elementary teachers to practice post-testing argumentation discussions in engineering design*. [Paper presentation.] American Society for Engineering Education Annual Conference and Exposition, Online. <https://doi.org/10.18260/1-2--34444>

Luna, M. J., & Sherin, M. G. (2017). Using a video club design to promote teacher attention to students' ideas in science. *Teaching and Teacher Education*, 66, 282-294. <https://doi.org/10.1016/j.tate.2017.04.019>

Marco-Bujosa, L., Gonzalez-Howard, M., McNeill, K., & Loper, S. (2017). Designing and using multimedia modules for teacher educators: Supporting teacher learning of scientific argumentation. *Innovations in Science Teacher Education*, 2(4), 1-16.

Martin, S. D., & Dismuke, S. (2018). Investigating differences in teacher practices through a complexity theory lens: The influence of teacher education. *Journal of Teacher Education*, 69(1), 22-39. <https://doi.org/10.1177/0022487117702573>

Masters, H. (2020). Using teaching rehearsals to prepare preservice teachers for explanation-driven science instruction. *Journal of Science Teacher Education*, 31(4), 414-434. <https://doi.org/10.1080/1046560X.2020.1712047>

McDonald, M., Kazemi, E., & Kavanagh, 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. <https://doi.org/10.1177/0022487113493807>

McGraw, A., & Davis, R. (2017). Mentoring for pre-service teachers and the use of inquiry-oriented feedback. *International Journal of Mentoring*

and Coaching in Education, 6(1), 50-63. <https://doi.org/10.1108/IJMCE-03-2016-0023>

McLeod, R. H., Kim, S., & Resua, K. A. (2019). The effects of coaching with video and email feedback on preservice teachers' use of recommended practices. *Topics in Early Childhood Special Education, 38*(4), 192–203. <https://doi.org/10.1177/0271121418763531>

McManus, S. (2008). *Attributes of effective formative assessment*. Council of Chief State School Officers: Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards.

McNeill, K. L., Katsh-Singer, R., González-Howard, M., & Loper, S. (2016). Factors impacting teachers' argumentation instruction in their science classrooms. *International Journal of Science Education, 38*(12), 2026-2046. <https://doi.org/10.1080/09500693.2016.1221547>

Mikeska, J. N., & Howell, H. (2020). Simulations as practice-based spaces to support elementary science teachers in learning how to facilitate argumentation-focused science discussions. *Journal of Research in Science Teaching, 57*(9), 1356-1399. <https://doi.org/10.1002/tea.21659>

Mikeska, J. N., & Howell, H. (2021, March 30). *Pushing the boundaries of practice-based teacher education: How can online simulated classrooms be used productively to support STEM teacher learning?* <https://aaas-arise.org/2021/03/30/pushing-the-boundaries-of-practice-based-teacher-education-how-can-online-simulated-classrooms-be-used-productively-to-support-stem-teacher-learning/>

Mikeska, J. N., & Lottero-Perdue, P. S. (2022). How preservice and in-service elementary teachers engage student avatars in scientific argumentation within a simulated classroom environment. *Science Education, 106*(4), 980-1009. <https://doi.org/10.1002/sce.21726>

Mikeska, J.N., Howell, H., & Kinsey, D. (2022). *Exploring the potential of simulated teaching experiences as approximations of practice within elementary teacher education: Do simulated teaching experiences impact elementary preservice teachers' ability to facilitate discussions in mathematics and science?* [Manuscript submitted for publication]. K-12 Center for Teaching, Learning, and Assessment, ETS.

Mikeska, J. N., Howell, H., & Straub, C. (2019). Using performance tasks within simulated environments to assess teachers' ability to engage in coordinated, accumulated, and dynamic (CAD) competencies. *International Journal of Testing, 19*(2), 128-147. <https://doi.org/10.1080/15305058.2018.1551223>

Mikeska, J. N., Webb, J., Bondurant, L., Kwon, M., Imasiku, L., Domjan, H., & Howell, H. (2021). Using and adapting simulated teaching experiences to support preservice teacher learning. In P.H. Bull & G.C. Patterson (Eds.), *Redefining teacher education and teacher preparation*

*programs in the post-COVID-19 era* (pp. 46-78) IGI Global. <https://doi.org/10.4018/978-1-7998-8298-5>

Moses, I. E. (1952). Nonparametric statistics for psychological research. *Psychological Bulletin*, 49(2), 122-143. <https://doi.org/10.1037/h0056813>

National Research Council. (2007). *Taking science to school: Learning and teaching science in grade K-8*. National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. National Academies Press.

Osborne, J., Simon, S., Christodoulou, A., Howell-Richardson, C., & Richardson, K. (2013). Learning to argue: A study of four schools and their attempt to develop the use of argumentation as a common instructional practice and its impact on students. *Journal of Research in Science Teaching*, 50(3), 315-347. <https://doi.org/10.1002/tea.21073>

Percy, M. M., & Troyan, F. J. (2017). Making transparent the challenges of developing a practice-based pedagogy of teacher education. *Teaching and Teacher Education*, 61, 26-36. <https://doi.org/10.1016/j.tate.2016.10.005>

Pella, S. (2015). Pedagogical reasoning and action: Affordances of practice-based teacher professional development. *Teacher Education Quarterly*, 42(3), 81-101. <https://www.jstor.org/stable/teaceducquar.42.3.81>

Prilop, C. N., Weber, K. E., & Kleinknecht, M. (2020). Effects of digital video-based feedback environments on pre-service teachers' feedback competence. *Computers in Human Behavior*, 102, 120-131. <https://doi.org/https://doi.org/10.1016/j.chb.2019.08.011>

Radloff, J., & Guzey, S. (2017). Investigating changes in preservice teachers' conceptions of STEM education following video analysis and reflection. *School Science and Mathematics*, 117(3-4), 158-167. <https://doi.org/10.1111/ssm.12218>

Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning and Instruction*, 27, 63-73. <https://doi.org/10.1016/j.learninstruc.2013.03.002>

Scheeler, M. C., Ruhl, K. L., & McAfee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, 27(4), 396-407. <https://doi.org/10.1177/088840640402700407>

Schreier, M. (2014). Qualitative content analysis. In U. Flick (Ed.), *The SAGE Handbook of qualitative data analysis* (pp. 170-183). SAGE Publications Ltd.



Sleep, L., Boerst, T., & Ball, D. (2007, March 21-24). *Learning to do the work of teaching in a practice-based methods course*. [Paper presentation.] National Council of Teachers of Mathematics Annual Meeting, Atlanta, GA.

Straub, C., Dieker, L., Hynes, M., & Hughes, C. (2015). Using virtual rehearsal in TLE TeachLivE™ mixed reality classroom simulator to determine the effects on the performance of science teachers: A follow-up study (Year 2). In *Proceedings of 3rd National TLE TeachLivE™ Conference* (pp. 49-109).

Tekcumru-Kisa, M., Stein, M. K., & Coker, R. (2018). Teachers' learning to facilitate high-level student thinking: Impact of a video-based professional development. *Journal of Research in Science Teaching*, 55(4), 479–502. <https://doi.org/10.1002/tea.21427>

Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85, 475-511. <https://doi.org/10.3102/0034654314564881>

Van der Kleij, F. M., & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability*, 33(2), 345-373. <https://doi.org/10.1007/s11092-020-09331-x>

Wiggins, G. (2012). Seven keys to effective feedback. *Educational Leadership*, 70(1), 10-16. <https://doi.org/10.3102/0034654308320292>

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, 1, 80-83. [https://doi.org/10.1007/978-1-4612-4380-9\\_16](https://doi.org/10.1007/978-1-4612-4380-9_16)

William, D. (2013). Feedback and instructional correctives. In J. H. McMillan (Ed.), *Handbook of research on classroom assessment* (pp. 197-214). Sage.

Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education*, 96, 878-903. <https://doi.org/10.1002/sce.21027>

Zeichner, K. (2010). Rethinking the connections between campus courses and field experiences in college-and university-based teacher education. *Journal of Teacher Education*, 61(1-2), 89-99. <https://doi.org/10.1177/0022487109347671>

*Contemporary Issues in Technology and Teacher Education* is an online journal. All text, tables, and figures in the print version of this article are exact representations of the original. However, the original article may also include video and audio files, which can be accessed online at <http://www.citejournal.org>

## Appendix A

### Description of the Five Rubric Features

Rubric Feature	Description of Feature	Key Scoring Indicators
1: Attending to Student Ideas	This dimension involves being responsive to students and focuses on making sure that the discussion is grounded in students' ideas and that all students are engaged in some meaningful component of the discussion.	1a. All of the key ideas that appear in the student work samples are incorporated into the discussion. (Incorporates Key Ideas Represented in Students' Prework) 1b. All student voices are heard in some nontrivial way, and the teacher probes for key ideas from all students. (Elicits Substantive Contributions) 1c. Each of the relevant student ideas is attended to by the teacher and made a part of the discussion. (Makes Use of Students' Ideas)
2: Facilitating a Coherent and Connected Discussion	This dimension involves the degree to which the teacher is able to shape a coherent discussion and focuses in particular on building and connecting ideas toward a learning goal.	2a. The teacher organizes the discussion in a coherent manner. (Overall Coherence of the Discussion) 2b. The teacher tracks the shape of the lesson in a way that is apparent to the students. (Making the Content Storyline Apparent to Students) 2c. The teacher strives to make connections and to link ideas in substantive ways. (Connections)
3: Promoting Peer Interactions	This dimension involves the ways that teachers strive to facilitate the discussion so that students are responsible for explaining key ideas.	3a. The students and the teacher share responsibility for the discussion. (Peer Interaction) 3b. The students are encouraged to engage with one another's ideas. (Engagement with Others' Ideas) 3c. The students generate ideas to move the discussion forward. (Ideas Come from Students)
4: Developing Students' Conceptual Understanding	This dimension involves the extent to which the teacher and students are involved in the evaluation of ideas that are put forth during the discussion.	4a. Who evaluates the accuracy and validity of ideas. (Who Evaluates the Ideas) 4b. Presence, absence, and acknowledgement of content errors. (Teacher Errors)
5: Engaging Students in Argumentation	This dimension involves the degree to which students engage in argumentation.	A focus on building consensus and/or evaluating opposing ideas. (Disciplinary Argumentation)

## Appendix B

### *Codes and Frequencies for Teachers' Attention to Their Main Takeaway/Learning From the Feedback*

<b>Code</b>	<b>Describes a takeaway or learning...</b>	<b>All Participants (n = 15)</b>	<b>Scoring Level Feedback Group (n = 8)</b>	<b>Specific Feedback Group (n = 7)</b>
Attending to Student Ideas (Feature 1)	... linked to being responsive to students and focuses on making sure that the discussion is grounded in students' ideas and that all students are engaged in some meaningful component of the discussion.	2 (13%)	2 (25%)	0 (0%)
Facilitating a Coherent Discussion (Feature 2)	... linked to shaping a coherent discussion and focuses in particular on building and connecting ideas toward a learning goal	6 (40%)	6 (75%)	0 (0%)
Promoting Peer Interactions (Feature 3)	... linked to facilitating a discussion so that students are responsible for explaining key ideas and interacting with one another.	7 (47%)	3 (38%)	4 (57%)
Developing Students' Conceptual Understanding (Feature 4)	... linked to the extent to which teacher and students are involved in the evaluation of ideas that are put forth during the discussion and/or focusing on helping students to build their conceptual understanding.	9 (60%)	5 (63%)	4 (57%)
Engaging Students in Argumentation (Feature 5)	... linked to engaging students in argumentation.	7 (47%)	4 (50%)	3 (43%)
Generic Feature	... linked to the rubric features in the formative feedback statements but does not identify details about what specifically was learned in regard to one or more features.	4 (27%)	4 (50%)	0 (0%)
Use Time Differently	... linked to how the teacher could have or did use the discussion time differently.	3 (20%)	1 (13%)	2 (29%)
Strengths	... linked to a strength from their performance identified in the formative feedback report.	5 (33%)	3 (38%)	2 (29%)
Areas for Improvement	... linked to an area of improvement from their performance identified in the formative feedback report.	14 (93%)	8 (100%)	6 (86%)
Vague/Not Related	... that is too vague to categorize or is not related to the question.	4 (27%)	1 (13%)	3 (43%)
Other	... that is not one of the previous codes.	1 (7%)	0 (0%)	1 (14%)

## Appendix C

### *Codes and Frequencies for Teachers' Use of Feedback*

<b>Code</b>	<b>Subcode</b>	<b>Describes use of formative feedback...</b>	<b>All Participants (n = 15)</b>	<b>Scoring Level Feedback Group (n = 8)</b>	<b>Specific Feedback Group (n = 7)</b>
Approach	Teaching Moves	...linked to adjusting or modifying the teaching moves or activities used for second Shoreline performance.	13 (87%)	6 (75%)	7 (100%)
	Discussion Goals	...linked to adjusting or modifying the discussion goals for the second Shoreline performance.	3 (20%)	3 (38%)	0 (0%)
	Discussion Structure	...linked to adjusting or modifying the structure of the discussion for the second Shoreline performance.	6 (40%)	3 (38%)	3 (43%)
	Outline/Notes	...linked to adjusting or modifying the notes or outline used for the second Shoreline performance.	8 (53%)	4 (50%)	4 (57%)
Substance	Attending to Student Ideas (Feature 1)	...linked to being responsive to students and focuses on making sure that the discussion is grounded in students' ideas and that all students are engaged in some meaningful component of the discussion.	10 (67%)	4 (50%)	6 (86%)
	Facilitating a Coherent Discussion (Feature 2)	...linked to shaping a coherent discussion and focuses in particular on building and connecting ideas toward a learning goal.	10 (67%)	7 (88%)	3 (43%)
	Promoting Peer Interactions (Feature 3)	...linked to facilitating a discussion so that students are responsible for explaining key ideas and interacting with one another.	11 (73%)	4 (50%)	7 (100%)
	Developing Students' Conceptual Understanding (Feature 4)	...linked to the extent to which teacher and students are involved in the evaluation of ideas that are put forth during the discussion and/or focusing on helping students to build their conceptual understanding.	11 (73%)	6 (75%)	5 (71%)
	Engaging Students in Argumentation (Feature 5)	...linked to engaging students in argumentation.	8 (53%)	5 (63%)	3 (43%)
	Use Time Differently	...linked to how the teacher could have or did use the discussion time differently.	3 (20%)	2 (25%)	1 (14%)
	Vague/Not Related	...that is too vague to categorize or not related to the question.	1 (7%)	1 (13%)	0 (0%)
	Other	...that is not one of the previous codes.	1 (7%)	1 (13%)	0 (0%)