Examining the Connection between Preservice Teachers' Discussion Performance in a Mixed Reality Teaching Simulation with their Self-Reported Goals and Success in Facilitating Discussions

Abstract

With the growing use of mixed reality teaching simulations in teacher education there is a need for researchers to examine how preservice teacher (PST) learning can be supported when using these simulations. To address this gap the current study explores how 47 PSTs used an online teaching simulation to facilitate a discussion focused on argumentation with five student avatars in the *Mursion™* mixed reality simulated classroom environment. We assessed PSTs' performance in the simulation using rubric-level scores assigned by trained raters and then compared the scores to PSTs' survey responses completed after their discussion asking them to self-report their goals for the discussion, how successful they thought they were across five dimensions of facilitating high-quality, argumentation-focused discussions, and their overall perceptions of the mixed reality teaching simulation. Findings suggest that PSTs' understanding of the discussion task's learning goals somewhat predicted their success in facilitating the discussion and that PSTs' self-assessment of their performance was not always consistent with raters' evaluation of the PSTs' performance. In particular, self-assessment was found to be most consistent with raters' evaluations for those PSTs with higher rater-assigned scores and least consistent for those with lower rater-assigned scores. The implications of these findings are as follows: (1) researchers should be cautious in relying on PST self-report of success when engaging in mixed reality teaching simulations, particularly because low performance may be obscured, (2) teacher educators should be aware that reliance on self-report from PSTs likely obscures the need for additional support for exactly those PSTs who need it most, and (3) the field, therefore, should expand efforts to measure PSTs' performance when using mixed reality teaching simulations.

Key words: simulated teaching, mixed reality simulations, self-assessment, discussions, practice-based teacher education

1. Introduction

Teachers play a crucial role in impacting student outcomes (Chamberlain, 2013; Chetty et al., 2014). To be able to engage in high-quality instruction that positively impacts student outcomes, preservice teachers (PSTs) need opportunities to learn in and from their practice. Typically, these learning opportunities occur by having PSTs work with mentor teachers as they engage in student teaching assignments at local schools (Goldhaber et al., 2020; Jenset et al., 2019; McDonald et al., 2013). However, research has shown that quality mentors are not always easy to come by (Greenberg et al., 2011). Student teaching also requires PSTs to engage in multiple instructional practices at one time as they teach entire lessons, which precludes them from focusing on specific instructional skills, such as eliciting student thinking or facilitating small group discussions (Mikeska, Howell, & Kinsey, 2023). Additionally, student teaching tends to happen in isolation from PSTs' coursework within methods courses (Darling-Hammond, 2006), which makes it challenging for PSTs to connect the teaching strategies they observe or use in classrooms to the theoretical ideas they learn about in their courses. This disconnect has come under scrutiny, as the disconnect makes it difficult for PSTs to make sufficient progress in learning when, how, and why to deploy specific teaching practices effectively to support student outcomes (Forzani, 2014; McDonald et al., 2013; Peercy & Troyan, 2017). To address some of these limitations of student teaching, during the last decade researchers and teacher educators (TEs) have turned towards using innovative tools, such as mixed reality teaching simulations, to broaden the learning opportunities available to PSTs when ascertaining how to engage in key teaching practices.

Mixed reality teaching simulations provide opportunities for PSTs to practice their ability to enact specific instructional skills in more controlled settings and allow them to analyze their performance through observation of artifacts related to their performance (e.g., video record, transcript) (Anthony et al., 2015; DeMink-Carthew et al., 2017) enabling them to deconstruct their practice and encourage reflection on their teaching strategies. These kinds of simulations also provide others, such as TEs, their peers, mentor teachers, or trained raters, with the ability to watch the PST's performance and provide them with feedback about the strengths and areas for growth evident in their instruction. Such feedback can be useful to support PSTs' reflection and improvement. However, while reflection and evaluation of practice is an important aspect of being an effective teacher, research has suggested that PSTs can struggle to accurately assess their practice (Clipa et al., 2011; Luke et al., 2021).

The growing use of mixed reality teaching simulations has prompted researchers to investigate how these simulations work to support PST learning. To date, most research in this area tends to use self-reports from PSTs as outcome measures to assess the usefulness of these tools to support PST learning – although there are a few notable exceptions (Kaka et al., 2021; Howell & Mikeska, 2021; Mikeska & Howell, 2020; 2021a; Poultsakis et al., 2021; Thompson et al., 2019). To better understand the potential of simulated tools to support PST learning, the field needs to conduct studies that examine: (a) how PSTs understand the key teaching goals they should address when engaged in these simulations, (b) the nature of PSTs' self-reflections when they engage in mixed reality teaching simulations and how these self-reflections compare to objective measures, such as raters' assessment of PSTs' instruction, and (c) PSTs' perceptions of the mixed reality teaching simulations. Such research would allow the field to better understand how PST learning can be supported when using mixed reality teaching simulations, the nature of the evidence that is best used to examine their learning when using mixed reality teaching simulations, and effective practices for incorporating mixed reality teaching simulations into teacher education settings. In the present study, we focus on three specific gaps in the field's understanding of mixed reality teaching simulation use. First, we explore the degree to which PSTs understand the goal of the task they are given and how this relates to their success. Second, we explore the extent to which PST self-report is likely to be useful and accurate either to support TEs' instructional choices or to draw conclusions about the learning that transpires when using mixed reality simulations. And finally, we explore whether PSTs see their performance as a reasonably accurate measure of their skills. To address these gaps, we investigate the following research questions (RQs):

- **RQ1a:** To what extent do PSTs' reported discussion goals for a mixed reality teaching simulation task align with the provided discussion task learning goals?
- **RQ1b:** How does alignment of PSTs' reported discussion goals for a mixed reality teaching simulation task connect to their discussion performance scores assigned by trained raters?
- **RQ2:** How, if at all, is PSTs' reported success in a mixed reality teaching simulation task consistent with their discussion performance rubric scores assigned by trained raters?
- **RQ3:** How did PSTs perceive the mixed reality teaching simulation?

In what follows, we begin by providing background about the use of mixed reality teaching simulations and the use of self-reflection to support teacher learning. We then turn to a description of the study's methodology, including the study's sample and data collection and analysis methods. We end by sharing the study's findings and implications in terms of using mixed reality teaching simulations as tools to support PST learning of key teaching practices and the nature of PSTs' self-reflections, in comparison to raters' assessment of PSTs' instruction, when using these tools.

2. Theoretical Framework

The current study is grounded in a theory of practice-based teacher education (Ball & Forzani, 2009; Forzani, 2014; Grossman, Compton, et al., 2009; Lampert, 2010; Zeichner, 2012). This theory suggests that teacher learning occurs through opportunities to learn in and from their practice. Three key pedagogies of practice are used to provide these practice-based learning opportunities: decompositions, representations, and approximations (Grossman, Compton, et al., 2009). Grossman, Compton, and colleagues (2009) propose employing representations to cultivate the ways in which PSTs perceive and comprehend professional practice. Additionally, they advocate for the use of decompositions to enhance PSTs' understanding and implementation of the various elements of teaching. This involves breaking down the practice of teaching into fundamental components, such as concentrating on aspects of lesson planning. However, the current study focuses on providing opportunities for PSTs to practice enacting the actual work of teaching via approximations of practice – "opportunities to rehearse and enact discrete components of complex practice in settings of reduced complexity" (Grossman, Hammerness, & McDonald, 2009, p. 283). Approximations of practice are essential for PST learning as they afford PSTs the opportunity to make errors while trying out novel or challenging aspects of teaching practice and allow associated stakeholders, such as TEs, to provide feedback on the teaching practice (Grossman, Compton, et al., 2009).

Two common approximations of practice are peer rehearsals and digital teaching simulations. Peer rehearsals often happen face-to-face and involve PSTs engaging in the role of a teacher while one or more of their peers play the role of a K-12 student (Davis & Boerst, 2014; Ghousseini, 2017; Kazemi et al., 2016). On the other hand, digital teaching simulations use technology to allow PSTs to interact with their peers via an online chat (Thompson et al., 2022; Wang et al., 2021) or to interact with virtual students (Kamhi-Stein et al., 2020; Lottero-Perdue et al., 2023; Mikeska, Shekell, et al., 2022; Straub, 2018). The use of digital teaching simulations as approximations of practice is described in the next section.

3. Literature Review

3.1. Digital Teaching Simulations in Teacher Education

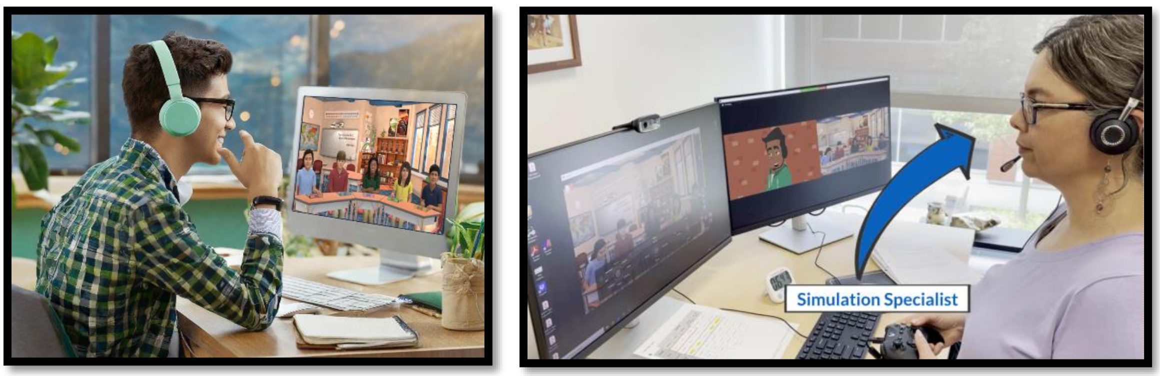
During the last decade, digital teaching simulations have become more frequently used as tools to support PST learning (Bondie & Dede, 2021; Cohen et al., 2020; Dalinger et al., 2020; Dieker et al., 2014; Ersozlu et al., 2021). Across the varied simulations used, the focus tends to be on helping PSTs refine their instructional practice and learn how to engage in specific teaching skills, such as eliciting student thinking, facilitating classroom discussions, promoting a positive classroom environment, or working with students who are second-language learners or who have disabilities (Bondie et al., 2021; Mikeska, Howell, Dieker, & Hynes, 2021; Regalla et al., 2016). In addition to supporting learning of key instructional practices, simulations have been used to develop PSTs' content knowledge, pedagogical content knowledge, preparedness to teach, self-efficacy, and beliefs about high-quality instruction and student learning (Kaufman & Ireland, 2019; Mikeska, Cross Francis, et al., 2023; Theelen et al., 2019). Despite the variation in learning outcomes targeted, all simulations provide PSTs with practice teaching opportunities in controlled and safe settings of reduced complexity that are low risk for both PSTs and students – although the specific approaches used vary across simulated environments.

One approach is to have PSTs engage in text-based chats with each other, using an online interface where each PST's identity and image is hidden (Reich, 2022; Wang et al., 2021). This approach is similar to face-to-face peer rehearsals, as PSTs are engaged in direct interaction with each other – with one PST acting as the teacher and the other acting as a K-12 student. For example, the Eliciting Learner Knowledge simulation is an online text-based chat where one PST, acting as the teacher, attempts to elicit the other PST's thinking, who is acting as an elementary student, about a science or mathematics topic (Mikeska, Shekell, et al., 2022; Thompson et al., 2022). The goal in the Eliciting Learning Knowledge simulation is to probe the student's thinking and to understand the student's prior conceptions or experiences related to the science or mathematics topic. The aim is not to alter the student's thinking or address any misconceptions they may have.

Another approach is for PSTs to respond in writing to prompts within an online interface that shows various agents or avatars responding in writing to the teacher or each other, with the teacher selecting next steps from a menu of pre-selected options (Deale & Pastore, 2014; Gibson, 2013; McPherson et al., 2011; Meritt et al., 2015). These online simulated environments provide PSTs with access to a small group or classroom of virtual students where they can engage in an aspect of the work of teaching, albeit in a reduced complexity setting. For example, simSchool is an online simulated environment where PSTs can log on, select the number of students in their class (individual, small group, or whole class), make instructional decisions, and observe students' physical, emotional, and academic responses to the decisions they make (Christensen et al., 2011; Gibson, 2007). SimSchool has been used across various studies to help PSTs and in-service teachers learn how to engage in effective teaching strategies, with findings across studies indicating that PSTs report increased confidence, attitudes, and teaching skills after using simSchool (Badiee & Kaufman, 2014; Collum et al., 2018; Hopper, 2018).

Most recently, one of the more common approaches is the use of a fully interactive online classroom consisting of five student avatars who can respond in real-time to a teacher's questions or comments and to other student avatars' ideas (Lee et al., 2023; Mikeska, Howell, & Kinsey, 2022, 2023; Straub et al., 2015). Figure 1 shows an example of one such classroom, *Mursion*^{TM}'s upper elementary mixed reality simulated classroom. In these types of mixed reality simulations, the student avatars are controlled on the backend by a person called a simulation specialist who has completed in-depth training to know how to respond as each of the student avatars during the simulated teaching session. Teachers can interact in real-time with the five student avatars and the avatars can respond to the teacher's verbal prompts and questions as well as interact directly with one another –like teacher and student verbal interactions in a real classroom. Usually, the teacher's interactions with the student avatars are video recorded and the video record is provided to the teacher afterward for review and reflection. To date, *Mursion™*'s mixed reality simulated classrooms has been used as a practice space for PSTs and in-service teachers to learn how to facilitate mathematics and science discussions, to elicit student thinking, to work with second language learners and students with disabilities, to facilitate productive parent conferences, and to enact successful classroom management (Bondie & Dede, 2021; Dawson & Lignugaris/Kraft, 2017; Dieker et al., 2019; O'Callaghan & Piro, 2016; Vince Garland & Garland, 2020). Previous research has indicated that PSTs and in-service teachers perceive these types of mixed reality simulations as authentic to the work of teaching (Mikeska & Howell, 2021a; Thompson et al., 2019).

Figure 1
Preservice Teacher (left) and Simulation Specialist (right) Interacting in the Mursion™ Mixed Reality Simulated Classroom



Note. Images reproduced with permission of *Mursion™* and Towson University.

Studies in this area tend to focus on PSTs' perceptions of the authenticity and usability of these mixed reality teaching simulations and determining what PSTs learn from their interactions in these simulations; it is rare for studies to investigate PSTs' understanding of their goals when engaging in the mixed reality simulation. In terms of authenticity, research suggests that PSTs and in-service teachers tend to perceive the simulations as useful contexts to practice aspects of instruction and the interactions with student avatars are reasonable approximations for how real students respond and behave, although arguments have been made that conceptualizations of authenticity in the contexts of mixed reality teaching simulations are currently undertheorized (Howell & Mikeska, 2021). In terms of learning outcomes, the use of mixed reality teaching simulations has shown to improve multiple aspects of PST learning including their ability to facilitate argumentation-focused discussions (Mikeska, Howell, & Kinsey, 2023), their perceptions about the importance of and confidence with facilitating controversial discussions in their classrooms (Kaka et al., 2021), their classroom management skills and teaching dispositions (Kaufman & Ireland, 2016), and how well they develop questioning strategies and feel sympathy towards students (Wang et al., 2021). Yet, to date, many studies tend to rely on PSTs' self-reflections on these varied outcome measures, rather than objective measures generated from external review of PSTs' actual performance in the mixed reality teaching simulations. While PSTs' self-reflections hold much potential, there have also been concerns and challenges about such reflections cited in previous research, which is where we turn next.

3.2. The Use and Validity of Self-Reflection in Teacher Education

Reflection is a common practice that has been used for decades in teacher education to support PST learning (Gay & Kirkland, 2003; Goodman, 1984; Pedro, 2005; Schön, 1983). It can be used by PSTs to reflect on their instructional practice in the moment – for example, as they interact with students during student teaching or as they engage in a mixed reality teaching simulation with student avatars – although most reflection tends to occur either before instruction or after instruction, as PSTs reflect on what went well and how they could adjust their instruction next time (Schön, 1983). As such, reflection can be variably characterized as something that occurs in action, as a cognitive activity, or as a combination of the two (Bengtsson, 1995). PSTs' reflective practices can focus on a myriad of instructional features and factors, ranging from their ability to manage the classroom, support student content-focused learning, improve student self-efficacy or confidence, or engage students in sense-making. In addition, PSTs' reflective practices can vary in both form and context – for example, PSTs can engage in self-reflection via participation in a video club with other PSTs, as part of written reflections in teaching portfolios, through structured conversations with mentor teachers, or individually by recording their observations as the view and reflect on a video record of their own teaching practice (Ajayi, 2016; Davis, 2006; Delandshere & Arens, 2003; Harford & MacRuairc, 2008; Kong, 2010; Xiao & Tobin, 2018).

Research has suggested that PSTs tend to view self-reflection in a positive light and believe that engaging in reflective practices benefits their learning and developing practice. For example, Erdemir and Yeşilçınar's (2021) recent study found that PSTs held positive views about engaging in self-reflection as they valued it for enhancing their motivation, raising their awareness of their teaching performance, and providing a way for them to critically analyze their teaching practice. Others have noted how self-reflection can allow PSTs to identify specific areas where they need to improve and develop their pedagogical knowledge (Johnson & Cotterman, 2015; Önal, 2019; Snead & Freiberg, 2019). In a similar vein, studies have shown that using structured reflection tools, such as rubrics or frameworks, to evaluate PSTs' teaching can be perceived as useful to improve their teaching competencies and knowledge (Davis & McDonald, 2019; van Diggelen et al., 2013).

Despite these frequently cited benefits to engaging PSTs in self-reflection, other studies have suggested that these reflections may vary for PSTs and novices in terms of the depth and accuracy of their self-reflections on their own and other's instructional practices (Blaik Hourani, 2013; Poom-Valickis & Mathews, 2013; Saric & Steh, 2017). For example, Goldman and Grimbeek (2015) qualitatively analyzed elementary PSTs' written reflections during student teaching and found that they were much more adept at describing and making judgments about specific teaching issues or challenges and relating, or connecting, these issues to themselves and their professional roles. However, their study indicated that the PSTs' written reflections rarely showed evidence of them engaging in complex pedagogical reasoning, reconstructing a plan of action, or metacognitively examining various factors or influences at play. Likewise, Selkrig and Keamy (2015) point to challenges for both PSTs and TEs alike when supporting PSTs to engage in critical reflection about their own and other's instructional practice.

Taken together, these studies suggest both benefits and potential drawbacks to PSTs' self-reflections. In particular, PSTs may experience significant challenges in being able to: (a) identify their strengths and areas for improvement, (b) engage in in-depth pedagogical reasoning about the teaching challenges they experience, or (c) sufficiently determine possible next steps to address challenges they are experiencing. Despite these previously cited concerns with the validity of PSTs' self-reflections, much of the research examining the use of mixed reality teaching simulations relies upon such self-reflections to report on learning outcomes. It is unclear to what extent this extensive reliance on PSTs' self-reflections is warranted or potentially problematic when using mixed reality teaching simulations to examine and support their pedagogical development. To our knowledge, this study is the first of its kind to investigate and compare the relationship between PST self-reflections and more objective measures of PST performance in the context of mixed reality teaching simulations.

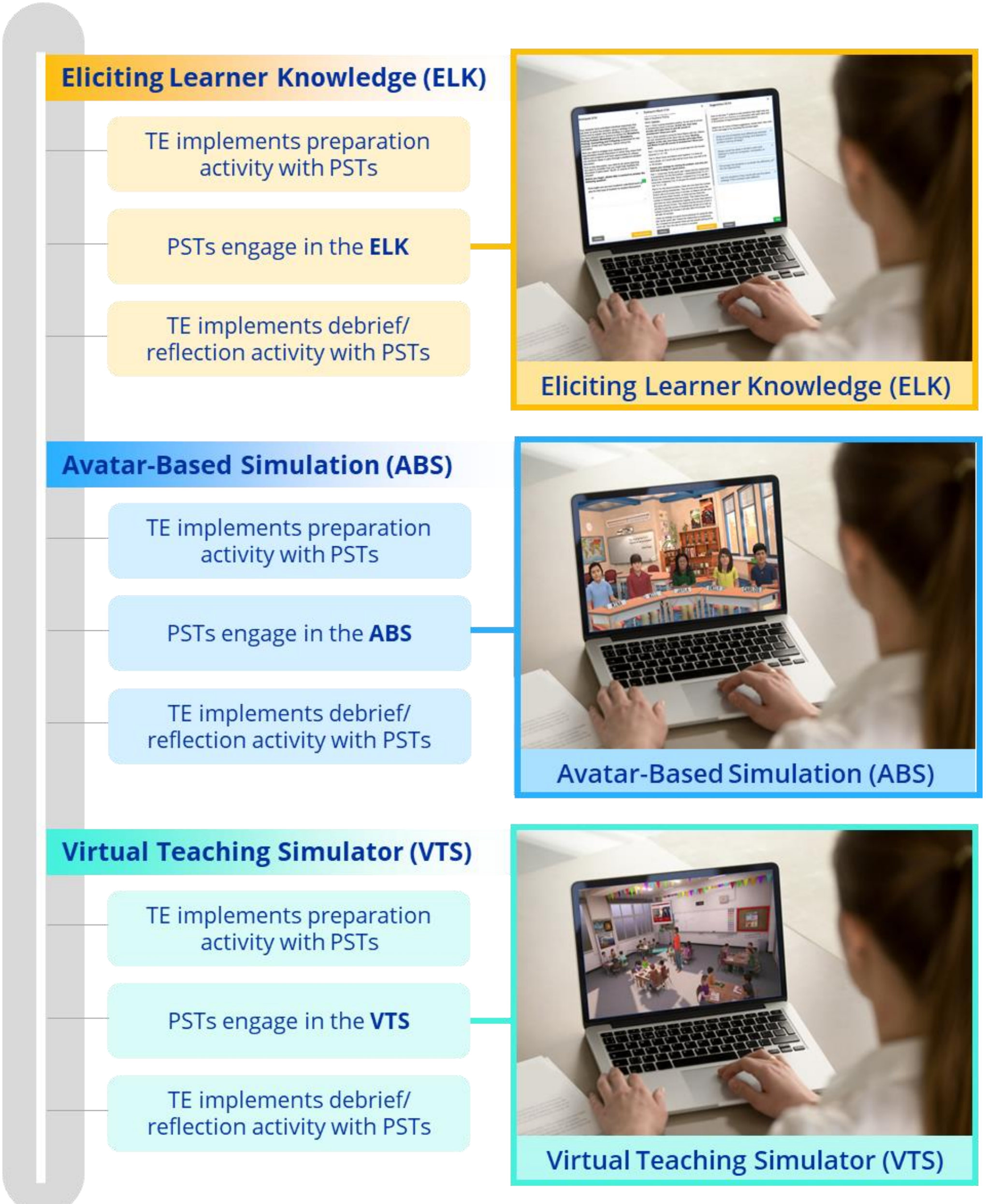
4. Material and Methods

4.1. Study Context

The current study is part of a larger research project where our team supported elementary TEs with the implementation of an online practice suite (OPS) of three teaching simulations within their mathematics and science teaching methods courses across one semester. The three simulations focused on enhancing PSTs' ability to facilitate argumentation-focused discussions. The PSTs enrolled in the mathematics methods courses completed the OPS centered on the topic of comparing fractions whereas those enrolled in the science methods courses completed the OPS centered on the topic of conservation of matter. For each of the three simulations the TE implemented a microcycle in which they first engaged their PSTs in one or more preparation activities prior to the teaching simulation and then implemented one or more debrief or reflection activities after the simulation (see Figure 2). TEs were not strictly instructed on how to engage their PSTs in the OPS, however TEs were provided with a guidebook that included information about each of the simulations and had access to an electronic bank of resources to support their instructional design of the preparation and debrief activities. These resources included articles, examples of activities they could implement, and video and transcript examples of previous PSTs completing the simulations.

The first simulation, Eliciting Learner Knowledge (ELK), involves a pair of PSTs role playing as either a student or a teacher and engaging in a text-based chat about the student's ideas. The second simulation, the Avatar-Based Simulation (ABS), is the focus of the present study and involves PSTs facilitating a small group discussion with five student avatars in the *Mursion*^{TM} simulated classroom environment. The third simulation, the Virtual Teaching Simulator (VTS), involves PSTs embodying a teacher avatar in a class of 24 student avatars. Please refer to Figure 2 or Mikeska, Shekell, et al. (2022) for a detailed description of the larger research project and the two additional teaching simulations used in the OPS.

Figure 2
Online Practice Suite Simulations



4.2. *Mursion*™ Mixed Reality Simulated Classroom Environment

The current study only focused on the ABS instead of all three simulations in the OPS as ABS was the only simulation that had PST self-assessment data and scoring from trained raters at the time of analysis. The ABS required PSTs to complete a 20-minute discussion with five virtual student avatars in the *Mursion*™ upper elementary simulated classroom environment. The five student avatars are controlled by a human-in-the-loop, also referred to as a simulation specialist. Once a PST joined their discussion session they were first greeted by an adult avatar (also portrayed by the simulation specialist) that was there to make sure the technology was working properly. The PST then conducted a short warm-up task with the student avatars prior to their 20-minute discussion; this warm-up task involved collecting the student avatars' lunch orders for that day. The purpose of the warm-up task was to allow the PST time to acclimate to the mixed reality simulated classroom before facilitating the mathematics or science discussion. Once the warm-up task was completed the PST then facilitated an up to 20-

minute argumentation-focused discussion on the given mathematics or science topic. Once the discussion was complete the adult avatar returned to ask the PST a few brief reflection questions about how their discussion went. Within 48 hours of completing their mixed reality teaching simulation the PST and their TE were able to access and review a recording of the PST's discussion video via the *Mursion™* portal.

The five student avatars in the *Mursion™* mixed reality simulated classroom environment are controlled by a simulation specialist who uses specialized technology to respond as each of the student avatars. PSTs never see the simulation specialist and only see the five student avatars on their screen as if they were on a Zoom meeting with the student avatars [see Figure 1 or OPS (2023a; 2023b) for further detail]. Different voice morph settings are applied for each of the student avatars to allow one simulation specialist to sound like five different upper elementary students. *Mursion™* provides the simulation specialists with a different vocal profile, vocal signature, and vocal energy/pace for each of the student avatars. For this project three simulation specialists were thoroughly trained by project staff on how to respond as each of the five student avatars based on that particular student's understanding or misunderstanding of the task content (comparing fractions and conservation of matter) and their previous experiences in class. While three simulation specialists were trained for the project, only one simulation specialist participated during each of the discussion sessions.

The *Mursion™* mixed reality platform uses a proprietary combination of human and artificial intelligence (AI) technology with highly realistic renderings of hundreds of computer-generated avatars and environments (Mursion, n.d.-a) that are developed using motion capture and gaming technology (Mursion, n.d.-b). The web-based platform is accessed via all standard web enabled devices with minimal internet bandwidth required (less than 4 mbps) and is compatible with Windows and Mac operating systems (Mursion, n.d.-a). *Mursion™* addresses universal demands by having global delivery and multi-language capabilities, meeting Web Content Accessibility Guidelines (WCAG) Level 2.1 AA with assistive software for accessibility needs and guaranteeing security and privacy of users through compliance with General Data Protection Regulation (GDPR) and Service Organization Control Type 2 (SOC 2) (Mursion, n.d.-a).

4.3. Sample

Participants included 47 elementary PSTs enrolled in one of the five mathematics or science teaching methods courses that implemented the OPS during the spring semester in 2022. Initially five TEs teaching methods courses at five public universities and colleges in the Southwestern, Midwestern, Northeastern, Southeastern, and Mid-Atlantic regions of the United States were recruited. Our PST sample was then comprised of the PSTs within the TEs' courses who agreed to participate in the study's research component. The current study included two mathematics methods courses and three science methods courses with 12 PSTS in the first math course, eight PSTs in the second math course, 11 PSTs in the first science course, seven PSTs in the second science course, and nine PSTs in the third science course. PSTs engaged in the OPS microcycles alongside the PSTs enrolled in their same methods course. The PSTs did not interact with the PSTs enrolled in the other methods courses involved in the current study, however, TEs did collaborate with one another in planning how to introduce the simulations in their classes.

All 47 participants consented to participate in the study's research components, completed their ABS discussion, and completed the ABS task survey after their discussion session. Participants were typically in their third or fourth year of a bachelor's program (74%) and most of them identified as female (96%) and White (79%). Participants also identified as Hispanic/Latino (19%), Black or African American (6%), American Indian or Alaskan Native (4%), and Asian or Asian American (2%). Most participants (94%) had no previous experience using a mixed reality teaching simulation; however, more than three quarters of participants (78%) had previously received feedback on their ability to lead a discussion with students.

4.4. Data Sources and Collection

From January to May of 2022, the three teaching simulations used in the OPS were integrated into five elementary mathematics and science methods courses. The second simulation, ABS, required PSTs to facilitate a 20-minute mathematics (Howell, Mikeska, Tierney, et al., 2021) or science (Mikeska, Howell, Orlandi, et al., 2021) discussion using the *Mursion*™ upper elementary simulated classroom. One week prior to the simulated discussion session, PSTs received a task packet to use in their preparation. The task packet included the student learning goal for the discussion, information about the previous class activities the student avatars engaged in before the discussion, copies of the student avatars' written work, and details about specific aspects the PSTs should notice about the student avatars' work. After the mixed reality simulated discussion was complete and the PSTs reviewed their discussion recording, the PSTs completed the ABS task survey to capture perceptions of their experience participating in the ABS activity. PSTs' responses to the ABS task survey were used for the current study.

The research team also hired six trained raters to apply a four-level rubric developed in a previous project and later refined (GO Discuss Project, 2021) to assess PSTs' ability to engage in five dimensions, or features, of facilitating high-quality, argumentation-focused discussions:

1) attending to students' ideas (i.e., making sure every student's voice is heard and that all students' ideas are valued) (Dimension 1),
2) facilitating a coherent and connected discussion (i.e., facilitating a discussion that makes sense and feels organized and purposeful) (Dimension 2),
3) encouraging student-to-student interactions (i.e., getting students to engage in discussion with one another) (Dimension 3),
4) developing students' conceptual understanding (i.e., supporting students in developing correct content understanding during the discussion) (Dimension 4), and
5) engaging students in argumentation (i.e., encouraging students to engage in argument construction, justification, critique, and/or building consensus) (Dimension 5).

Each PST video was assigned five individual dimension scores across the four levels of the rubric (level 1 = beginning practice, level 2 = developing practice, level 3 = well-prepared practice, level 4 = commendable practice). Each dimension was comprised of three indicators that also received scores from raters. For example, Dimension 4 is focused on developing students' conceptual understanding and is comprised of the following three indicators: (1) whether the students or teacher evaluated the accuracy and validity of ideas in the discussion, (2) the presence, absence, and acknowledgement of content errors made by the teacher, and (3) whether critical student misunderstandings related to the learning goal were addressed. Raters completed bias training, training on how to include evidence and effective justifications to explain the indicator and dimension scores they assigned, and specific training for each of the five dimensions and the indicators that comprise each dimension. Raters then practiced scoring a validity video performance previously scored and reconciled by project staff. One project staff member reviewed the raters' scoring for the validity video and provided feedback to each rater on how their scores and evidence aligned with the project staff's reconciled scores and evidence. If the rater met all training benchmarks, they were then able to start scoring the PSTs' discussions using the video and transcript of the performance. Twenty five percent of PST performances were double scored by two raters with adequate overall interrater reliability calculated to be .753 one-way average measures intraclass coefficient (ICC).

4.5. Analysis

A mixed methods analysis approach was implemented for the current study in which the authors used qualitative analysis to analyze the PSTs' open-ended survey responses and quantitative analysis to identify patterns across their self-reported goals and the raters' scores. Detailed descriptions of each analysis approach are included below.

4.5.1. RQ1: PST-reported Discussion Goal Alignment

To address RQ1a, we examined the alignment of PSTs' reported discussion goals for the mixed reality simulated discussion to the learning goal provided to PSTs in the task packet one week prior to their discussion. Specifically, we drew on two sources to inform the analysis. The first was the PSTs' responses from the ABS task survey question: "In your own words, what were the goals for the discussion focused on argumentation you facilitated in the simulated classroom? In other words, what did you want the students to do or learn in the simulated discussion?". The second, described in the next paragraph, was the learning goals as stated in the task packet PSTs received prior to leading their discussion. It is important to note that this is more than simply a test of whether teachers recall what was given in the course packet – teachers, including teachers in training, must always synthesize stated goals, make sense of them in their own words, and often layer their own understandings over what is given as they plan and execute a lesson. By asking PSTs to report in their own words, we elicit information about what they understood the goal to be rather than what we told them it should be.

The learning goal for Emily's One Less Method task (comparing fractions) in mathematics is as follows, "Students will construct an argument about whether a method for generating a fraction in between two given fractions is worth remembering so that it can be used in the future." (Howell, Mikeska, Tierney, et al., 2021, p. 3). The learning goal for the Making Lemonade task (conservation of matter) in science is as follows, "Students will construct an argument, using evidence and reasoning, about whether the amount of matter changes when ingredients are mixed together to make lemonade. They will build consensus about which observations to use as evidence that matter is conserved during physical changes." (Mikeska, Howell, Orlandi, et al., 2021, p. 3).

An evidence inventory was developed to indicate if a PSTs' self-reported goal focused on at least one aspect of argumentation and one aspect of task content, each of which was represented in the mathematics or science task goal. For instance, accurate aspects of task content would be present if a science PST mentioned a goal about having the student avatars determine which observations to use as evidence that matter is conserved during physical changes or if a math PST mentioned a goal about whether a method for generating a fraction in between two given fractions is worth remembering so that it can be used in the future. Assessing for accurate aspects of argumentation involved examining how the PST-reported goal involved engaging the student avatars in achieving or addressing the content pieces previously mentioned. For example, accurate argumentation would be present if a science PST mentioned the student avatars coming to consensus or if a math PST mentioned the student avatars formulating an argument.

We then calculated a three-level alignment classification of strong/partial/no alignment based on the presence of evidence from both categories of the evidence inventory. See Table 1 for examples of PST-reported learning goals in each of the three alignment classifications. Strong alignment was achieved when a PSTs' discussion goal included an accurate focus on one or more aspects of argumentation and one or more aspects of the task content. Partial alignment occurred when the PSTs' discussion goal included an accurate focus on either an argumentation or task content aspect. Finally, no alignment occurred if the PSTs' discussion goal did not include an accurate focus on an argumentation or task content aspect represented in the mathematics or science discussion task. Two researchers used the evidence inventory and alignment classification to code 27 science and 20 mathematics PSTs' responses. Adequate rater agreement and interrater reliability—calculated using one-way average measures ICC—were achieved when coding both science (93% exact rater agreement, .925 ICC) and mathematics (97% exact rater agreement, .968 ICC) responses. Additionally, Cohen's weighted kappa was run, and almost perfect agreement (McHugh, 2012) was achieved when coding science, κ_w= .860 (95% CI, .771 to .948), $p < .001$, and math responses, κ_w = .938 (95% CI, .868 to 1.007), $p < .001$.

Table 1
PST Learning Goal Examples by Alignment Classification

Alignment Classification	Math Examples	Science Examples
Strong Alignment	*I wanted them to understand the points that other students made about the One Less Method and be able to respond accordingly. I wanted students to discuss and debate the use of keeping the method in their toolbox.* (PST 5012)	*The goals were to get students to come to an agreement as to whether or not matter was conserved using evidence to support their conclusion.* (PST 1006)
Partial Alignment (Argumentation)	*I wanted students to learn and understand the different approaches to solving the problem. I wanted them to be able to critique each other's ideas respectfully and critically.* (PST 4004)	*I wanted the students to have a scientific argument where they would share their findings and support them with evidence and reasoning. If others had different findings, they could share the details to come to an overall consensus.* (PST 2006)
Partial Alignment (Task Content)	*I wanted the students to decide if they should keep Emily's one less method in their toolbox* (PST 5011)	*The students' learning goal was to find out if the amount of matter changed when mixing ingredients to make lemonade.* (PST 3003)
No Alignment	*I wanted students to find the one less method most helpful* (PST 4005)	*I wanted them to learn about how matter and volume were different.* (PST 1015)

Further analysis was conducted for RQ1b by comparing PSTs' mean performance scores assigned by trained raters across the five dimensions of facilitating argumentation-focused discussions when PSTs were grouped by their alignment classification category (strong alignment, partial alignment, or no alignment). We examined mean scores by goal-alignment categorization, attending closely to whether the directionality was as expected, and ran correlations to determine the relationship between alignment group and mean scores across the five dimensions as well as the overall score across dimensions.

4.5.2. RQ2: PST-reported Success

To address the second research question, we examined the consistency of PSTs' self-assessment of their success on the five dimensions of facilitating argumentation-focused discussions to the raters' scores on each dimension using the four-level scoring rubric. For the self-assessment aspect, we used PSTs' responses from the ABS task survey question asking PSTs to indicate how successful they were in implementing the five elements of a high-quality discussion focused on argumentation during their ABS session using a four-point Likert scale (not at all successful, minimally successful, somewhat successful, very successful).

For purposes of this study, we examined the general consistency of the PSTs' self-assessment Likert scale responses to the raters' rubric scores. To assess this general consistency, PSTs' self-assessment on the lower end of the scale who reported they were not at all successful or minimally successful in a dimension were combined into a "Low Self-Assessment" category, while PSTs who reported they were somewhat successful or very successful were grouped into a "High Self-Assessment" category.

Furthermore, the rater scores assigned to the PSTs' performances for each dimension were grouped similarly in which beginning practice (level 1) and developing practice (level 2) were combined into a "Low Rater Score" category and well-prepared practice (level 3) and commendable practice (level 4) were grouped into a "High Rater Score" category. We decided to group the two lowest levels of the rubric and the two highest levels of the rubric together, as more nuanced groupings would have resulted in minimal sample sizes across groups for comparison purposes.

"Consistency" occurred when the PSTs' self-assessment and the raters' score were both in the low categories or both in the high categories. For example, consistency would be present if a PST indicated they were very successful (High Self-Assessment) for Dimension 1 (attending to students' ideas) on the ABS task survey and the rater assigned that PST's performance in Dimension 1 at the well-prepared practice (level 3) (High Rater Score). "Inconsistency" would be present if a PST indicated they were very successful (High Self-Assessment) for Dimension 1 and a rater assigned that PST's performance in Dimension 1 at the developing practice level (level 2) (Low Rater Score). Inconsistency included two categories to indicate (1) when a PSTs' self-assessment was higher than the rater score (i.e., High Self-Assessment and Low Rater Score) and (2) when a PSTs' self-assessment was lower than the rater score (i.e., Low Self-Assessment and High Rater Score). Frequencies across the groups were identified and compared to determine if there were any patterns by dimension or overall across these groups.

4.5.3. RQ3: PST Perceptions of the Mixed Reality Simulation

To address the third research question, we examined frequencies of PSTs' responses to the ABS task survey in which PSTs were asked to indicate their agreement on a four-point Likert-type scale (disagree, somewhat disagree, somewhat agree, agree) with five statements regarding their perceptions of the mixed reality simulation. If a PST disagreed or somewhat disagreed with the statement "My performance in the simulated classroom environment accurately reflects my ability to facilitate classroom discussions with real students", the PST was prompted with an open response follow-up question asking them to explain why. We used qualitative content analysis (Schreier, 2012) and a codebook developed from a previous study (Mikeska & Howell, 2021b) to group responses into themes identified across PST responses (codes and descriptions included in Table 5).

5. Results and Discussion

5.1. RQ1: PST-reported Discussion Goal Alignment

5.1.1. RQ1a

In total, eight (17%) of the 47 responses in which PSTs described the task goals in their own words were categorized as strongly aligned to the task goals specified in the mathematics or science discussion task, indicating that the PST's response showed signs of both critical components of the stated learning goal – an accurate description of at least one element of argumentation and an accurate reference to the content focus of the task. Findings showed that most responses (29; 62% of total responses) were categorized as partially aligned, meaning that the PST's response contained evidence of one but not both components. Of these, 12 (26% of total responses) included the argumentation component but not the content and 17 (36% of total responses) included content but not argumentation. The remaining responses (10; 21% of total responses) were categorized as no alignment as they did not include reference to either of the two components.

5.1.2. RQ1b

Our next step was to compare the alignment of the learning goal to the rater-assigned scores on each dimension and overall. This analysis was aimed at seeking evidence to support the logical hypothesis that a PST who shows stronger understanding of the goal of the task might be more likely to be successful in meeting that goal, meaning PSTs in the strong alignment group would have higher scores than the partially aligned group and the no alignment group, and PSTs in the partially aligned group would have higher scores than the no alignment group. Means and standard deviation of scores on each dimension and overall, broken down by the PST goal alignment categorization are shown in Table 2. We hypothesized that PSTs with stronger alignment to the learning goal would be more likely to have higher rater scores. Overall, the trend of mean scores was as we expected in which PSTs in the strong alignment group consistently scored higher on each dimension and overall compared to those in the partial and no alignment groups, although in some cases the difference was quite small such as in Dimension 5 (engaging students in argumentation). We also anticipated that PSTs in the partially aligned group would consistently score higher on each dimension and overall compared to those in the no alignment group, however the results were less consistent and more variable with only Dimensions 2 (facilitating a coherent and connected discussion) and 4 (developing students' conceptual understanding) and the overall scores following this pattern.

Table 2

Comparing PSTs' Dimension Score Means by Discussion Goal Alignment Group (N = 47 PSTs)

Discussion Goal Alignment Group	Dimension 1 (attending to students' ideas) M (SD)	Dimension 2 (facilitating a coherent and connected discussion) M (SD)	Dimension 3 (encouraging student-to-student interactions) M (SD)	Dimension 4 (developing students' conceptual understanding) M (SD)	Dimension 5 (engaging students in arg.) M (SD)	Overall Across Dimensions M (SD)
Strong Alignment (*n* = 8)	3.44 (.90)	2.63 (1.30)	2.88 (1.13)	3.06 (.78)	2.31 (.96)	14.31 (4.54)
Partial Alignment (*n* = 29)	3.03 (.83)	2.36 (.80)	2.03 (.52)	2.79 (.88)	2.28 (.66)	12.50 (2.84)
No Alignment (*n* = 10)	3.10 (.74)	2.15 (.47)	2.25 (.72)	2.55 (.76)	2.30 (.48)	12.35 (2.38)
Average Scores Across Participants (*N* =47)	3.12 (.82)	2.36 (.85)	2.22 (.74)	2.79 (.84)	2.29 (.67)	12.78 (3.11)

Six correlations were run to determine the relationship between alignment group and mean scores across the five dimensions as well as the overall score across dimensions. No statistically significant correlations were determined and all Pearson's correlation coefficients were less than .30, therefore were classified as weak correlations (Akoglu, 2018). Despite the lack of statistical significance, the this relationship trends in the expected direction, with PSTs who show strong evidence of understanding the task goals to be more likely to achieve higher dimension-level and overall scores. This pattern suggests that PSTs are likely to do better if they have a clear understanding of what they are trying to address during the simulation, therefore it may be beneficial for TEs to spend time ensuring their PSTs have a strong understanding of the learning goal prior to leading a discussion in a mixed reality simulation.

5.2. RQ2: PST-reported Success

In examining the consistency between PST self-reported success and rater-assigned scores there were 235 units of analysis comparing a PST self-report success level and a dimension score across the five dimensions and 47 PSTs. Across these 235 units of analysis, findings showed that the PSTs' self-assessment was inconsistent with raters in 120 of 235 instances (51%), meaning PSTs' self-reported assessment of their success across these five dimensions were just as likely to be inconsistent with raters scores as they were to be consistent (49%) (see Table 3). Most instances of inconsistency (87%; 104 of 120 instances) occurred because of a high PST self-assessment and a low rater assessment, suggesting that PSTs who are struggling may have the greatest difficulty identifying their areas of improvement. Results showed no notable patterns when examining the number of dimensions each individual PST was consistent or inconsistent in, meaning if a PST was inconsistent on one dimension, they were not always inconsistent on the other dimensions with just over half of PSTs (26; 55% of PSTs) likely to be inconsistent on about half (two or three) of the dimensions.

Consistency of PSTs' self-assessment of their performance with the raters' scores occurred most often in Dimension 1 (attending to students' ideas) (72%), followed by Dimension 4 (developing students' conceptual understanding) (66%), Dimension 2 (facilitating a coherent and connected discussion) (38%), and Dimension 3 (encouraging student-to-student interactions) (34%), and Dimension 5 (engaging students in argumentation) (34%). Of note, when ordering the mean scores across all participants for each dimension from highest to lowest, with Dimensions 1, 4, 2, 5, and 3 having mean scores of 3.12, 2.79, 2.36, 2.29, and 2.22 (see bottom row of Table 2), respectively, the same pattern is present as when ordering dimensions by highest to lowest consistency of PSTs' self-assessment and rater scores. Meaning, the higher the mean score for the dimension, the higher chance of consistency between PSTs' self-assessment and rater scores for that dimension. This finding is consistent with previous research that has shown when comparing PSTs' assessment of their own performance to their TE's grades, PSTs with better performances had a more accurate self-assessment and more consistency to their TE's grades (Clipa et al., 2011).

Table 3

PST Self-Assessment and Rater Score Consistency

PST Self-Assessment and Rater Score Consistency	Dimension 1 (attending to students' ideas) n (%)	Dimension 2 (facilitating a coherent and connected discussion) n (%)	Dimension 3 (encouraging student-to-student interactions) n (%)	Dimension 4 (developing students' conceptual understanding) n (%)	Dimension 5 (engaging students in arg.) n (%)	Across Dimensions n (%)
Consistent	34 (72%)	18 (38%)	16 (34%)	31 (66%)	16 (34%)	115 (49%)
High Self-Assessment and High Rater Score	34 (100%)	16 (89%)	7 (44%)	27 (87%)	13 (81%)	97 (84%)
Low Self-Assessment and Low Rater Score	0 (0%)	2 (11%)	9 (56%)	4 (13%)	3 (19%)	18 (16%)
Inconsistent	13 (28%)	29 (62%)	31 (66%)	16 (34%)	31 (66%)	120 (51%)
High Self-Assessment and Low Rater Score	11 (85%)	26 (90%)	27 (87%)	13 (81%)	27 (87%)	104 (87%)

Low Self-Assessment and High Rater Score	2 (15%)	3 (10%)	4 (13%)	3 (19%)	4 (13%)	16 (13%)

Note. Percentages in the sub rows under Consistent are out of the number of instances that were consistent for that dimension. The same applies to the sub rows under Inconsistent.

Furthermore, consistency was more likely to occur when both the PST self-assessment and rater score were high. This pattern occurred most often in Dimension 1 (attending to students' ideas) with all instances of consistency (100%) being a result of high self-assessment and high rater scores. The only dimension that did not fit this pattern was Dimension 3 (encouraging student-to-student interactions) in which nearly half the instances of consistency occurred with high self-assessment and rater scores (44%) and the other half occurred with low self-assessment and rater scores (56%). Dimension 3 focuses on encouraging student-to-student interactions, and this finding could suggest that PSTs who struggle in this component are able to assess their weaknesses more accurately when compared to other components. For instance, it may be more obvious for PSTs to notice that the students are not interacting with one another whereas it may be more challenging for PSTs to know if they are facilitating a coherent and connected discussion (Dimension 2).

Altogether the results of the current study indicated that PSTs were only able to accurately assess their performance in 49% of instances, in comparison to raters' rubric scores. When PSTs were inconsistent with raters, they were almost always perceiving themselves as doing more positively on a particular aspect of argumentation-focused discussions. This finding suggests that TEs would be missing key information on their PSTs' areas of improvement if they were only using their PSTs' self-reflections from the mixed reality simulated teaching discussions. The main implication from this finding is that it is key to include multiple forms of feedback in teacher education to assist PSTs' and TEs' in the identification of important areas for improvement. Mixed reality simulations can be useful in teacher education to provide TEs with a platform to offer constructive feedback on PSTs' instructional skills, which allows for multiple iterations and refinement (Thompson et al., 2019). A well-executed feedback mechanism enables PSTs to scrutinize performance discrepancies, emphasizing both strengths and areas for improvement (Cornelius & Nagro, 2014; Sweigart et al., 2015). This, in turn, encourages PSTs to delve deeper into introspection and personal growth (Grossman & McDonald, 2008; Zhang & Cheng, 2011). Recognizing the pivotal role that reflection plays in the pedagogical development of PSTs, it is imperative to effectively utilize feedback to enhance this process. Fundamentally, feedback establishes the groundwork for reflection (Diana, 2013) and is widely regarded as the starting point of the reflective journey (Lichtenberger-Majzikné & Fischer, 2017). However, if one does not recognize that certain areas of instruction would benefit from additional practice and growth, it is difficult – if not impossible – to support improvement in those areas.

5.3. RQ3: PST Perceptions of the Mixed Reality Simulation

PSTs' agreement with five statements assessing their perceptions of the mixed reality simulation were mixed (see Table 4). First, frequency counts showed that around one third of the PSTs (31; 66%) disagreed or somewhat disagreed that working in a simulated classroom environment was unfamiliar, which made it difficult for them to do well on the ABS task. Furthermore, the majority of PSTs (42; 89%) agreed or somewhat agreed that the simulated classroom environment was easy to use. Additionally, more than half of PSTs (27; 57%) thought their performance in the simulated classroom environment did not accurately reflect their ability to facilitate classroom discussions with real students and many PSTs (31; 66%) thought they did not perform as well as they could have. However, most PSTs (41; 88%) thought their performance would improve by giving more opportunities to practice facilitating discussions in the simulated classroom environment, further supporting the need for these types of tools to be included in teacher education.

Table 4

PST Perceptions of the Mixed Reality Simulation (N = 47 PSTs)

Statement	Disagree n (%)	Somewhat Disagree n (%)	Somewhat Agree n (%)	Agree n (%)
Working in a simulated classroom environment was unfamiliar, which made it difficult for me to do well on this task.	13 (28%)	18 (38%)	9 (19%)	7 (15%)
I found the simulated classroom environment easy to use (even if the tasks themselves might have been difficult).	0 (0%)	5 (11%)	25 (53%)	17 (36%)
My performance in the simulated classroom environment accurately reflects my ability to facilitate classroom discussions with real students.	17 (36%)	10 (21%)	17 (36%)	3 (6%)
I did not perform as well as I could have.	4 (9%)	12 (26%)	21 (45%)	10 (21%)
Given more opportunities to practice facilitating discussions in the simulated classroom environment, my performance on this task would improve.	2 (4%)	4 (9%)	21 (45%)	20 (43%)

Twenty-seven PSTs were prompted to explain why they thought their performance in the simulated classroom environment did not accurately reflect their ability to facilitate classroom discussions with real students. Around one third of these 27 PSTs (17; 63%) indicated that the context of the simulation environment is dissimilar from what would normally be found in a real classroom (see Table 5). Responses referencing the context of the classroom environment included general explanations about using a virtual environment as well as more specific aspects such as the size of the group (i.e., five student avatars) and the time limit on the discussion. For example, as PST 4002 (math) explained, "Because it is through the screen with avatars, it is not as personal as it would be, teaching in a real classroom…". PSTs also commonly cited their own personal characteristics, such as their nervousness, as a reason why their performance failed to accurately reflect their ability (11; 41%). PST 3002 (science) cited both the simulation context and their own personal characteristics in their explanation: "…I became so stressed about the time limit, then when I realized my kiddos had not reached a consensus, I felt incredibly defeated". According to Cross Francis and colleagues (2020), elementary teachers undergo a complex range of emotions with fluctuating patterns during teaching, which can either foster productive teaching practices or exacerbate adverse physiological and cognitive responses hindering effective teaching for others.

Table 5

PSTs' Reasons for Why Their Simulated Discussion Performance Did Not Reflect Their Ability (N = 27 PSTs)

Code	Description *Describes an inaccurate performance due to…*	All PST Responses n (%)
Context	…the context of the simulation environment is dissimilar from what would normally be found in a real classroom.	17 (63%)
PST Personal Characteristics	…the PST's personal characteristics.	11 (41%)
Student Avatars	…the differences in how the student avatars respond and/or behave compared to real students.	5 (19%)
Inexperience with Simulated Classrooms	…the PST's lack of familiarity with using simulated classrooms.	2 (7%)

Lack of Preparation	…a lack of preparation. The lack of preparation could be either on the PST's or TE's part.	1 (4%)
Vague	…a reason that is too vague to categorize.	2 (7%)

Note. PSTs' responses could be assigned more than one code if applicable.

The differences in how the student avatars respond and behave compared to real students was also mentioned by PSTs as a reason for the inaccuracy of their performance (5; 19%). For instance, PST 3013 (science) mentioned what they perceived as a lack of authenticity in the student avatars' behavior:

> In the simulation, I am talking with adults acting as children. Their responses are catered to the questions that I am asking. In an actual classroom I will have to lead the discussion more because I will have more students that attempt to get us off topic or do not want to share at all. In an actual classroom, I will need to be using more of my teacher skill than what I am using in the simulation.

While we can appreciate PST 3013's concern regarding this perceived inauthenticity of the student avatars' behavior, previous research has suggested using teaching simulations in a structured and controlled manner can provide an opportunity for PSTs to experience aspects of classroom life, but with less complexity (Theelen et al., 2019). The aspect of classroom life that is the focus of the current study is helping PSTs learn how to facilitate argumentation-focused discussions, therefore we did not include any aspects of classroom management skills as part of the mixed reality simulated discussion experience to strategically reduce the complexity for PSTs. Finally, less common reasons for an inaccurate performance included the PSTs' lack of familiarity with using simulated classrooms (2; 7%) or a lack of preparation, either on the PST's or TE's part (1; 4%).

In general, these findings are similar to previous research that has shown mixed perceptions on the authenticity of online teaching simulations (Luke et al., 2021). Some research shows that PSTs do perceive simulations as authentic to the work of teaching, although certain aspects, such as limitations in what the avatars can do or how they behave, may be less authentic to what they experience in actual classrooms; it is important to note that the ability to control the simulations in these ways is what has been touted as one of the key benefits of using simulations as approximations to support PST learning (Mikeska & Howell, 2021a; Thompson et al., 2019). While unfamiliarity with the simulation is not a strong theme in the PST responses, many of the other codes (difference in context, nervousness, avatars seeming inauthentic, etc.) could well be linked to inexperience or to some degree ameliorated through increased experience with teaching simulations. Lack of experience in the simulator may contribute to a perceived inaccuracy in the PST's performance reflecting skill. While we have no way to measure the accuracy of that perception, a potential implication is that multiple exposures to such technology may be useful to increase the perceptions of accuracy of such mixed reality tools and provide TEs with useful information to make informed decisions about how to support PST learning.

5.4. Limitations

A limitation of the current study is that while our survey items are relatively straightforward and have been used in some form across multiple studies with no evidence of confusion by participants, we have not conducted reliability or validity studies on the surveys themselves. Furthermore, the sample size included in the current study may not have been sufficient enough to draw significant conclusions, specifically in RQ1.

6. Conclusions and Implications

Findings from the current study indicate that PSTs' understanding of the intended purpose, or aim, of the mixed reality teaching simulation – in this case, the student learning goals that were the intended focus of the mathematics or science discussion they facilitated – is somewhat a predictor of their success in facilitating argumentation-focused discussions. When comparing the mean scores assigned by trained raters for PSTs who had the strongest alignment to the learning goal and those with the weakest alignment to the learning goal there was a clear pattern of those with strongest alignment having higher mean scores. It is important to note that these findings do not suggest a clear causal relationship – as PSTs who showed better understanding of the learning goals may have likely also been the ones that had better teaching skills from the start. However, we argue that these results do suggest that knowing the learning goals that one is aiming for when engaging in a mixed reality teaching simulation is likely beneficial, as PSTs are unlikely to do well if they are unclear what they are trying to address during the simulation. One broader implication is that TEs might do well to focus preparation efforts on ensuring a robust understanding of the learning goals and might even consider assessing understanding of the task's goals as a measure of PST readiness to use the mixed reality simulation. Given the cost of implementing such mixed reality simulations, having a way to screen PSTs for readiness that is relatively easy for the TE to implement could be quite useful in ensuring that PSTs receive the best possible benefit from the experience by entering the mixed reality simulation ready to engage appropriately.

Furthermore, results showed that PSTs were only somewhat accurate in evaluating themselves on five dimensions of facilitating argumentation-focused discussion when compared to scores from trained raters. This finding supports previous research suggesting that PSTs are not always able to accurately assess their practice (Luke et al., 2021). However, we also found that the higher the PSTs' score for the dimension, the higher chance of consistency between PSTs' self-assessment and rater scores for that dimension. This finding is consistent with previous research showing PSTs with better performances had a more accurate self-assessment and more alignment to their TEs' grades (Clipa et al., 2011). The implication of these findings is that PST self-report is unlikely to be a reliable signal to TEs of which PSTs need more support. This is potentially consequential because most PSTs' only opportunity to practice teaching is during internships or field work, to which the TE often has little direct access. Therefore, it is common practice for TEs to rely on PST self-report on their success and learning from those experiences.

The previously stated conclusions suggest that a TE is as likely to have reliable information about a PSTs' performance simply by evaluating their understanding of the task goal as they are in relying on PST self-report, although neither is a strong predictor on their own. This conclusion suggests that TEs need direct access to tools that allow for objective assessment of PST skill and to access the information they need to make instructional decisions regarding their PSTs' learning. The field should therefore focus on providing tools that help measure PST skill directly. Mixed reality teaching simulations are one tool that can provide this, however one limitation of using these mixed reality tools and having objective measures of PST performance, such as scores from trained raters, is the expense required. Therefore, the field should also prioritize automation of these processes using artificial intelligence to the greatest extent reasonably possible.

The significance of this study lies in its contribution to the literatures on PST teacher learning and on use of simulations. Findings provide solid evidence that TEs should not rely exclusively on self-report of PSTs' success in teaching, and while this study focused on teaching in a mixed reality simulation, it suggests that self-report of other teaching might be equally suspect. It also suggests that the ubiquitous use of self-report measures to evaluate the success of mixed-reality simulations is potentially problematic on two fronts: first, it may be equally unreliable, and second, it may reflect only the adequacy of the participant's understanding of the task goals and not their skill in executing the task.

Statements on open data and ethics

This study has been reviewed and approved by the Institutional Review Board (IRB) at ETS. All procedures were performed in compliance with relevant laws and institutional guidelines, and the appropriate institutional committee(s) have approved them. All teacher educators and preservice teachers provided their written consent to participate in this research study. The data from the current study is included in a data repository. The data repository link and instructions can be provided upon request by sending an e-mail to the corresponding author.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ajayi, L. (2016, January). How intern teachers use classroom video for self-reflection on teaching. *The Educational Forum, 80*(1), 79-94. https://doi.org/10.1080/00131725.2015.1102365

Akoglu H. (2018). User's guide to correlation coefficients. *Turkish journal of emergency medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Anthony, G., Hunter, J., & Hunter, R. (2015). Prospective teachers development of adaptive expertise. *Teaching and teacher education*, *49*, 108-117. https://doi.org/10.1016/j.tate.2015.03.010

Badiee, F., & Kaufman, D. (2014). Effectiveness of an online simulation for teacher education. *Journal of Technology and Teacher Education*, *22*(2), 167-186. https://www.learntechlib.org/primary/d/45934/

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

Bengtsson, J. (1995). What is reflection? On reflection in the teaching profession and teacher education. *Teachers and teaching, 1*(1), 23-32. https://doi.org/10.1080/1354060950010103

Blaik Hourani, R. (2013). Pre-service teachers' reflection: Perception, preparedness and challenges. *Reflective Practice*, *14*(1), 12-30. https://doi.org/10.1080/14623943.2012.732947

Bondie, R., & Dede, C. (2021). Redefining and transforming field experiences in teacher preparation through personalized mixed reality simulations. In *What teacher educators should have learned from 2020* (pp. 229-242). Association for the Advancement of Computing in Education. https://www.learntechlib.org/p/219088/ebook_219088.pdf#page=230

Bondie, R., Mancenido, Z., & Dede, C. (2021). Interaction principles for digital puppeteering to promote teacher learning. *Journal of research on technology in education*, *53*(1), 107-123. https://doi.org/10.1080/15391523.2020.1823284

Chamberlain, G. E. (2013). Predictive effects of teachers and schools on test scores, college attendance, and earnings. *Proceedings of the National Academy of Sciences*, *110*(43), 17176-17182. https://doi.org/10.1073/pnas.1315746110

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633-2679. https://doi.org/10.1257/aer.104.9.2633

Christensen, R., Knezek, G., Tyler-Wood, T., & Gibson, D. (2011). SimSchool: An online dynamic simulator for enhancing teacher preparation. *International Journal of Learning Technology*, *6*(2), 201-220. https://doi.org/10.1504/IJLT.2011.042649

Clipa, O., Ignat, A. A., & Rusu, P. (2011). Relations of self-assessment accuracy with motivation level and metacognition abilities in pre-service teacher training. *Procedia-Social and Behavioral Sciences*, *30*, 883-888. https://doi.org/10.1016/j.sbspro.2011.10.171

Cohen, J., Wong, V., Krishnamachari, A., & Berlin, R. (2020). Teacher coaching in a simulated environment. *Educational evaluation and policy analysis*, *42*(2), 208-231. https://doi.org/10.3102/0162373720906217

Collum, D., Johnston, V., Blanks, B., & Delicath, T. (2018, March). A simulated classroom: A multi-university study evaluating simSchool's effectiveness on the educational needs of diverse learners. In *Society for Information Technology & Teacher Education International Conference* (pp. 1403-1412). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/p/182711/

Cornelius, K. E., & Nagro, S. A. (2014). Evaluating the evidence base of performance feedback in preservice special education teacher training. *Teacher Education and Special Education*, *37*(2), 133-146. https://doi.org/10.1177/0888406414521837

Cross Francis, D. I., Hong, J., Liu, J., Eker, A., Lloyd, K., Bharaj, P. K., & Jeon, M. (2020). The dominance of blended emotions: A qualitative study of elementary teachers' emotions related to mathematics teaching. *Frontiers in psychology*, *11*, 1865. https://doi.org/10.3389/fpsyg.2020.01865

Dalinger, T., Thomas, K. B., Stansberry, S., & Xiu, Y. (2020). A mixed reality simulation offers strategic practice for pre-service teachers. *Computers & Education*, *144*, 103696. https://doi.org/10.1016/j.compedu.2019.103696

Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of teacher education*, *57*(3), 300-314. https://doi.org/10.1177/0022487105285962

Davis, A., & McDonald, D. (2019). Teachers' reflections of instructional effectiveness: self-assessment through a standards-based appraisal process. *Reflective Practice, 20*(1), 125-141. https://doi.org/10.1080/14623943.2019.1569508

Davis, E. A. (2006). Characterizing productive reflection among preservice elementary teachers: seeing what matters. *Teaching and Teacher Education, 22*(3), 281–301. https://doi.org/10.1016/j.tate.2005.11.005

Davis, E. A., & Boerst, T. (2014). Designing elementary teacher education to prepare well-started beginners. *Ann Arbor, MI: TeachingWorks, University of Michigan School of Education*. https://www.teachingworks.org/images/files/TeachingWorks_Davis_Boerst_WorkingPapers_March_2014.pdf

Dawson, M. R., & Lignugaris/Kraft, B. (2017). Meaningful practice: Generalizing foundation teaching skills from TLE TeachLivE™ to the classroom. *Teacher Education and Special Education*, *40*(1), 26-50. https://doi.org/10.1177/0888406416664184

Deale, D., & Pastore, R. (2014). Evaluation of simSchool: An instructional simulation for pre-service teachers. *Computers in the Schools*, *31*(3), 197-219. https://doi.org/10.1080/07380569.2014.932650

Delandshere, G., & Arens, S. A. (2003). Examining the quality of the evidence in preservice teacher portfolios. *Journal of Teacher Education, 54*(1), 57-73. https://doi.org/10.1177/0022487102238658

DeMink-Carthew, J., Grove, R., & Peterson, M. (2017). The influence of the core practices movement on the teaching and perspectives of novice teacher educators. *Studying Teacher Education*, *13*(1), 87-104. https://doi.org/10.1080/17425964.2017.1289083

Diana, T. J. (2013). Microteaching revisited: Using technology to enhance the professional development of pre-service teachers. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *86*(4), 150-154. https://doi.org/10.1080/00098655.2013.790307

Dieker, L. A., Rodriguez, J. A., Lignugaris/Kraft, B., Hynes, M. C., & Hughes, C. E. (2014). The potential of simulated environments in teacher education: Current and future possibilities. *Teacher Education and Special Education*, *37*(1), 21-33. https://doi.org/10.1177/0888406413512683

Dieker, L. A., Straub, C., Hynes, M., Hughes, C. E., Bukathy, C., Bousfield, T., & Mrstik, S. (2019). Using virtual rehearsal in a simulator to impact the performance of science teachers. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, *11*(4), 1-20. https://doi.org/10.1177/0888406413512683

Erdemir, N., & Yeşilçınar, S. (2021). Reflective practices in micro teaching from the perspective of preservice teachers: Teacher feedback, peer feedback and self-reflection. *Reflective Practice*, *22*(6), 766-781. https://doi.org/10.1080/14623943.2021.1968818

Ersozlu, Z., Ledger, S., Ersozlu, A., Mayne, F., & Wildy, H. (2021). Mixed-reality learning environments in teacher education: An analysis of TeachLivE™ research. *Sage Open*, *11*(3). https://doi.org/10.1177/21582440211032155

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

Gay, G., & Kirkland, K. (2003). Developing cultural critical consciousness and self-reflection in preservice teacher education. *Theory into practice, 42*(3), 181-187. https://doi.org/10.1207/s15430421tip4203_3

Ghousseini, H. (2017). Rehearsals of teaching and opportunities to learn mathematical knowledge for teaching. *Cognition and Instruction*, *35*(3), 188-211. https://doi.org/10.1080/07370008.2017.1323903

Gibson, D. (2007). simSchool and the conceptual assessment framework. In *Games and simulations in online learning: Research and development frameworks* (pp. 308-322). IGI Global. https://doi.org/10.4018/978-1-59904-304-3.ch015

Gibson, D. (2013). Teacher education with simSchool. In *Teacher Education Programs and Online Learning Tools: Innovations in Teacher Preparation* (pp. 437-453). IGI Global. https://doi.org/10.4018/978-1-4666-1906-7.ch023

GO Discuss Project (2021). *Scoring*. Qualitative Data Repository. https://doi.org/10.5064/F6NJU10I

Goldhaber, D., Krieg, J., Naito, N., & Theobald, R. (2020). Making the most of student teaching: The importance of mentors and scope for change. *Education Finance and Policy*, *15*(3), 581-591. https://doi.org/10.1162/edfp_a_00305

Goldman, J. D., & Grimbeek, P. (2015). Pre-service primary school teachers' self-reflective competencies in their own teaching. *European Journal of Psychology of Education, 30*, 189-207. https://doi.org/10.1007/s10212-014-0231-8

Goodman, J. (1984). Reflection and teacher education: A case study and theoretical analysis. *Interchange, 15*(3), 9–26. https://psycnet.apa.org/doi/10.1007/BF01807939

Greenberg, J., Pomerance, L., & Walsh, K. (2011). Student teaching in the United States. *National Council on teacher quality*. https://eric.ed.gov/?id=ED521916

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal*, *45*(1), 184-205. https://doi.org/10.3102/0002831207312906

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

Harford, J., & MacRuairc, G. (2008). Engaging student teachers in meaningful reflective practice. *Teaching and teacher education, 24*(7), 1884-1892. https://doi.org/10.1016/j.tate.2008.02.010

Hopper, S. (2018). The heuristic sandbox: Developing teacher know-how through play in simSchool. *Journal of Interactive Learning Research*, *29*(1), 77-112. https://www.learntechlib.org/p/174847/

Howell, H., & Mikeska, J. N. (2021). Approximations of practice as a framework for understanding authenticity in simulations of teaching. *Journal of Research on Technology in Education*, *53*(1), 8-20. https://doi.org/10.1080/15391523.2020.1809033

Howell, H., Mikeska, J. N., Tierney, J., Baehr, B., & Lehman, P. (2021). *Conceptualization and development of a performance task for assessing and building elementary preservice teachers' ability to facilitate argumentation-focused discussions in mathematics: The fractions between task* (Research Memorandum No. RM-21-11). ETS. https://www.ets.org/Media/Research/pdf/RM-21-11.pdf

Jenset, I. S., Hammerness, K., & Klette, K. (2019). Talk about field placement within campus coursework: Connecting theory and practice in teacher education. *Scandinavian Journal of Educational Research*, *63*(4), 632-650. https://doi.org/10.1080/00313831.2017.1415968

Johnson, H. J., & Cotterman, M. E. (2015). Developing preservice teachers' knowledge of science teaching through video clubs. *Journal of Science Teacher Education, 26*(4), 393-417. https://doi.org/10.1007/s10972-015-9429-0

Kaka, S. J., Littenberg-Tobias, J., Kessner, T., Francis, A. T., Kennett, K., Marvez, G., & Reich, J. (2021). Digital simulations as approximations of practice: Preparing preservice teachers to facilitate whole-class discussions of controversial issues. *Journal of Technology and Teacher Education*, *29*(1), 67-90. https://www.learntechlib.org/primary/p/218711/

Kamhi-Stein, L. D., Lao, R. S., & Issagholian, N. (2020). The future is now: Implementing mixed-reality learning environments as a tool for language teacher preparation. *TESL-EJ*, *24*(3), n3. https://eric.ed.gov/?id=EJ1275837

Kaufman, D., & Ireland, A. (2016). Enhancing teacher education with simulations. *TechTrends*, *60*, 260-267. https://doi.org/10.1007/s11528-016-0049-0

Kaufman, D., & Ireland, A. (2019). Simulation as a strategy in teacher education. In *Oxford research encyclopedia of education*. https://doi.org/10.1093/acrefore/9780190264093.013.478

Kazemi, E., Ghousseini, H., Cunard, A., & Turrou, A. C. (2016). Getting inside rehearsals: Insights from teacher educators to support work on complex practice. *Journal of teacher education*, *67*(1), 18-31. https://doi.org/10.1177/0022487115615191

Kong, S. C. (2010). Using a web-enabled video system to support student–teachers' self-reflection in teaching practice. *Computers & Education, 55*(4), 1772-1782. https://doi.org/10.1016/j.compedu.2010.07.026

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

Lee, T. D., Lee, C., Newton, M., Vos, P., Gallagher, J., Dickerson, D., & Regenthal, C. (2023). Peer to peer vs. virtual rehearsal simulation rehearsal contexts: Elementary teacher candidates' scientific discourse skills explored. *Journal of Science Teacher Education*, 1-22. https://doi.org/10.1080/1046560X.2023.2181505

Lichtenberger-Majzikné, K., & Fischer, A. (2017). The role of feedback in developing reflective competence. *Practice and Theory in Systems of Education*, *12*(3), 119-127. https://www.irisro.org/ptse/4102KrisztinaLichtenbergerMajzikne-FischerAndrea.pdf

Lottero-Perdue, P. S., Cimino, K., & Brandeberry, J. (2023). Preservice teachers facilitating a discussion with elementary student avatars before facilitating it with real students. *Innovations in Science Teacher Education*, *8*(4). https://innovations.theaste.org/preservice-teachers-facilitating-a-discussion-with-elementary-student-avatars-before-facilitating-it-with-real-students/

Luke, S. E., Ford, D. J., Vaughn, S. M., & Fulchini-Scruggs, A. (2023). An online field experience using mixed reality virtual simulation. *Journal of Research on Technology in Education*, *55*(2), 324-343. https://doi.org/10.1080/15391523.2021.1962452

McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education: A call for a common language and collective activity. *Journal of teacher education*, *64*(5), 378-386. https://doi.org/10.1177/0022487113493807

McHugh M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, *22*(3), 276–282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

McPherson, R., Tyler-Wood, T., Ellison, A. M., & Peak, P. (2011). Using a computerized classroom simulation to prepare pre-service teachers. *Journal of Technology and Teacher Education*, *19*(1), 93-110. https://www.learntechlib.org/p/31438/

Meritt, J., Gibson, D., Christensen, R., & Knezek, G. (2015). Teacher training using interactive technologies: Performance and assessment in second life and simschool. In Isaías, P., Spector, J., Ifenthaler, D., Sampson, D. (Eds.), *E-learning systems, environments and approaches: Theory and implementation* (pp. 181-198). Springer, Cham. https://doi.org/10.1007/978-3-319-05825-2_13

Mikeska, J. N., & Howell, H. (2020). Simulations as practice-based spaces to support elementary 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. (2021a). Authenticity perceptions in virtual environments. *Information and Learning Sciences*, *122*(7/8), 480-502. https://doi.org/10.1108/ILS-10-2020-0234

Mikeska, J. N., & Howell, H. (2021b). *RAPID: Learning to teach during COVID-19: Leveraging simulated classrooms as practice-based spaces for preservice elementary teachers within online teacher education courses* (Award No. 2032179). National Science Foundation. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2032179&HistoricalAwards=false

Mikeska, J. N., Cross Francis, D. I., Lottero-Perdue, P. S., Park Rogers, M. A., Shekell, C., Bharaj, P. K., Howell, H., Maltese, A., Thompson, M. & Reich, J. (2023). Using a suite of online simulations to promote elementary preservice teachers' facilitation of argumentation-focused discussions in mathematics and science [Manuscript submitted for publication]. Research Institute, ETS. *Teaching and Teacher Education.* https://doi.org/10.17605/OSF.IO/TQWKV

Mikeska, J. N., Howell, H., & Kinsey, D. (2022). Examining the usability and viability of using a simulated classroom environment to prepare preservice science teachers during and after the COVID-19 pandemic. *Disciplinary and Interdisciplinary Science Education Research*, *4*(1), 1-20. https://doi.org/10.1186/s43031-022-00054-1

Mikeska, J. N., Howell, H., & Kinsey, D. (2023). Do simulated teaching experiences impact elementary preservice teachers' ability to facilitate argumentation-focused discussions in mathematics and science?. *Journal of Teacher Education*, *74*(5), 422-436. https://doi.org/10.1177/00224871221142842

Mikeska, J. N., Howell, H., Dieker, L., & Hynes, M. (2021). Understanding the role of simulations in K-12 mathematics and science teacher education: Outcomes from a teacher education simulation conference. *Contemporary Issues in Technology and Teacher Education, 21*(3), 781-812. https://citejournal.org/wp-content/uploads/2021/08/v21i3General.pdf

Mikeska, J. N., Howell, H., Orlandi, E., King, K., Lipari, M., & Simonelli, G. (2021). *Conceptualization and development of a performance task for assessing and building elementary preservice teachers' ability to facilitate argumentation-focused discussions in science: The conservation of matter task* (Research Memorandum No. RM-21-07). ETS. https://www.ets.org/Media/Research/pdf/RM-21-07.pdf

Mikeska, J. N., Shekell, C., Maltese, A. V., Reich, J., Thompson, M., Howell, H., Lottero-Perdue, P. S., & Rogers, M. P. (2022, April). Exploring the potential of an online suite of practice-based activities for supporting preservice elementary teachers in learning how to facilitate argumentation-focused discussions in mathematics and science. In *Society for Information Technology & Teacher Education International Conference* (pp. 1793-1803). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/noaccess/220984/

Mursion. (n.d.-a). *Our platform*. https://www.mursion.com/services/corporate/platform/

Mursion. (n.d.-b). *The neuroscience of practice*. https://www.mursion.com/why-it-works/

O'Callaghan, C., & Piro, J. (2016). Virtual simulations in a practice-based teacher education. *The Field Experience Journal*, *17*(1), 94-118. https://www.unco.edu/cebs/national-field-experience-conference/pdf/Volume_18.pdf#page=98

Önal, A. (2019). An exploratory study on pre-service teachers' reflective reports of their video-recorded microteaching. *Journal of Language and Linguistics Studies, 15*(3), 806–830. https://doi.org/10.17263/jlls.631520

Online Practice Suite (OPS). (2023a). *Avatar-Based Simulation (ABS): Elementary science* [Video]. The University of North Carolina at Chapel Hill School of Education. https://uncch.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=b83ff702-ad7a-47cd-bc77-b08100b970a5

Online Practice Suite (OPS). (2023b). *Avatar-Based Simulation (ABS): Elementary math* [Video]. The University of North Carolina at Chapel Hill School of Education. https://uncch.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=a5cea09a-8372-4617-83a0-b08100b99a3f

Pedro, J. (2005). Reflection in teacher education: Exploring pre-service teachers' meanings of reflective practice. *Reflective Practice*, *6*(1), 49–66. https://doi.org/10.1080/1462394042000326860

Peercy, 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

Poom-Valickis, K., & Mathews, S. (2013). Reflecting others and own practice: an analysis of novice teachers' reflection skills. *Reflective Practice, 14*(3), 420-434. https://doi.org/10.1080/14623943.2013.767237

Poultsakis, S., Papadakis, S., Kalogiannakis, M., & Psycharis, S. (2021). The management of digital learning objects of natural sciences and digital experiment simulation tools by teachers. *Advances in Mobile Learning Educational Research*, *1*(2), 58-71. https://doi.org/10.25082/AMLER.2021.02.002

Regalla, M., Hutchinson, C., Nutta, J., & Ashtari, N. (2016). Examining the impact of a simulation classroom experience on teacher candidates' sense of efficacy in communicating with English learners. *Journal of Technology and Teacher Education*, *24*(3), 337-367. https://www.learntechlib.org/p/171498/

Reich, J. (2022). Teaching drills: Advancing practice-based teacher education through short, low-stakes, high-frequency practice. *Journal of Technology and Teacher Education*, *30*(2), 217-228. https://www.learntechlib.org/p/221208/

Saric, M., & Steh, B. (2017). Critical reflection in the professional development of teachers: Challenges and possibilities. *CEPS journal, 7*(3), 67-85. https://doi.org/10.25656/01:14908

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Schreier, M. (2012). *Qualitative content analysis in practice*. Sage. https://us.sagepub.com/en-us/nam/book/qualitative-content-analysis-practice

Selkrig, M., & Keamy (R.) K. (2015). Promoting a willingness to wonder: moving from congenial to collegial conversations that encourage deep and critical reflection for teacher educators. *Teachers and Teaching, 2*1(4), 421–436. https://doi.org/10.1080/13540602.2014.969104

Snead, L. O., & Freiberg, H. J. (2019). Rethinking student teacher feedback: Using a self-assessment resource with student teachers. *Journal of Teacher Education, 70*(2), 155–168. https://doi.org/10.1177/0022487117734535

Straub, C. (2018). *Best in class leadership development: How virtual reality and avatars are changing the learning landscape*. Mursion, Inc.

https://www.denasamuels.com/wpcontent/uploads/2020/03/How-Virtual-Reality-and-Avatars-are-Changing-the-LearningLandscape.pdf

Straub, C., Dieker, L., Hynes, M., & Hughes, C. (2015, June). 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).

Sweigart, C. A., Landrum, T. J., & Pennington, R. C. (2015). The effect of real-time visual performance feedback on teacher feedback: A preliminary investigation. *Education and Treatment of Children*, *38*(4), 429-450. https://doi.org/10.1353/etc.2015.0024

Theelen, H., Van den Beemt, A., & den Brok, P. (2019). Classroom simulations in teacher education to support preservice teachers' interpersonal competence: A systematic literature review. *Computers & Education*, *129*, 14-26. https://doi.org/10.1016/j.compedu.2018.10.015

Thompson, M., Leonard, G., Mikeska, J. N., Lottero-Perdue, P. S., Maltese, A. V., Pereira, G., Hillaire, G., Waldron, R., Slama, R., & Reich, J. (2022). Eliciting learner knowledge: Enabling focused practice through an open-source online tool. Behavioral Sciences, 12(9), 324-341. https://doi.org/10.3390/bs12090324

Thompson, M., Owho-Ovuakporie, K., Robinson, K., Kim, Y. J., Slama, R., & Reich, J. (2019). Teacher Moments: A digital simulation for preservice teachers to approximate parent–teacher conversations. *Journal of Digital Learning in Teacher Education*, *35*(3), 144-164. https://doi.org/10.1080/21532974.2019.1587727

van Diggelen, M., den Brok, P., & Beijaard, D. (2013). Teachers' use of a self-assessment procedure: The role of criteria, standards, feedback and reflection. *Teachers and Teaching*, *19*(2), 115-134. https://doi.org/10.1080/13540602.2013.741834

Vince Garland, K., & Garland, D. (2020). TeachLivE™ and teach well: Simulations in teacher education. In E. Bradley (Ed.), *Games and Simulations in Teacher Education* (pp. 183-195). Springer International Publishing. https://doi.org/10.1007/978-3-030-44526-3_13

Wang, X., Thompson, M., Yang, K., Roy, D., Koedinger, K. R., Rose, C. P., & Reich, J. (2021). Practice-based teacher questioning strategy training with ELK: A role-playing simulation for eliciting learner knowledge. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1-27. https://doi.org/10.1145/3449125

Xiao, B., & Tobin, J. (2018). The use of video as a tool for reflection with preservice teachers. *Journal of Early Childhood Teacher Education, 39*(4), 328-345. https://doi.org/10.1080/10901027.2018.1516705

Zeichner, K. (2012). The turn once again toward practice-based teacher education. *Journal of teacher education*, *63*(5), 376-382. https://doi.org/10.1177/0022487112445789

Zhang, S., & Cheng, Q. (2011). Learning to teach through a practicum-based microteaching model. *Action in Teacher Education*, *33*(4), 343-358. https://doi.org/10.1080/01626620.2011.620523