


Article

Impact of Data-Driven Feedback and Coaching on Preservice Teachers' Questioning Skills for Higher-Order Thinking within a Mixed-Reality Simulation Environment

Wes J. DeSantis ¹, Marcia A. B. Delcourt ^{1,*}, Bruce M. Shore ²  and Jacob C. Greenwood ³

¹ Department of Education and Educational Psychology, School of Professional Studies, Western Connecticut State University, Danbury, CT 06810, USA; desantisw@wcsu.edu

² Department of Educational and Counselling Psychology, Faculty of Education, McGill University, Montréal, QC H3A 0G4, Canada; bruce.m.shore@mcgill.ca

³ Ridgefield Public Schools, Ridgefield, CT 06877, USA; jgreenwood@ridgefieldps.net

* Correspondence: delcourtm@wcsu.edu

Abstract: In this exploratory study, constructs related to self-efficacy theory were embedded in a treatment to improve the use of higher-order thinking (HOT) questioning skills in candidates enrolled in a teaching-methods course enhanced by mixed-reality simulations (MRS). The problem of designing an effective feedback and coaching model to improve the delivery of HOT questions in 10 min lessons was addressed. Thirty undergraduates were asked to incorporate HOT questions into each of the three lessons presented during a 15-week semester. Treatment candidates received individual data-driven feedback and coaching that included tailored guidance provided at regular intervals throughout the term. Quantitative analyses indicated that there was no significant difference in self-efficacy between conditions and that treatment group members posed significantly more HOT questions in their lessons (effect size = 1.26) than their non-treatment peers. An optimal ratio of two knowledge/comprehension to one HOT question in a 10 min period was proposed and three criteria for high-quality HOT questions are presented. Interviews revealed that those who participated in the treatment were more likely to recognize improvements in their self-efficacy, lesson planning, and performance than comparison group members. Data-driven feedback and coaching also provided candidates with opportunities for reflection.

Keywords: data-driven feedback and coaching; mixed-reality simulations; preservice teacher education; reflection; self-efficacy



Citation: DeSantis, W.J.; Delcourt, M.A.B.; Shore, B.M.; Greenwood, J.C. Impact of Data-Driven Feedback and Coaching on Preservice Teachers' Questioning Skills for Higher-Order Thinking within a Mixed-Reality Simulation Environment. *Educ. Sci.* **2023**, *13*, 596. <https://doi.org/10.3390/educsci13060596>

Academic Editor: Maria B. Peterson-Ahmad

Received: 12 April 2023

Revised: 25 May 2023

Accepted: 6 June 2023

Published: 11 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

An essential competency for a preservice teacher is to “develop high-level thinking, reasoning, and problem-solving skills” in their students [1]. These skills promote higher-order thinking (HOT), a crucial cognitive activity which teachers can systematically insert into classroom lessons by using questioning techniques [2–4]. HOT competencies reflect the top four levels of Bloom’s taxonomy of cognition: application, analysis, synthesis, and evaluation [5], or the four highest categories of the revised hierarchy of thinking: applying, analyzing, evaluating, and creating [6]. Entry-level skills for these taxonomies include knowledge and comprehension or K/C [5], or remembering and understanding [6].

Implementing high-leverage practices, Refs. [1,7,8] such as HOT skills, into classroom lessons requires training and is facilitated by appropriate feedback from experts [9–11]. Essentially, feedback is the confirmation of being correct or incorrect about a performance [10]. Receiving feedback is important for the growth of preservice teachers [12], however, ambiguous feedback, i.e., not clearly specifying what has or has not been achieved regarding a goal, can lower future performance [10]. For preservice teachers to comprehensively address specific areas in need of improvement, Kraft and Blazer [13] also asserted that

direct feedback should be provided immediately following the delivery of a lesson or after reviewing a digitally recorded one. Overall, “feedback is more effective the more information it contains” related to the target task [11]. Accurately implemented feedback enhances self-efficacy [14–16] because progress can be gauged as a candidate rehearses a specific competency, making periodic adjustments to achieve the desired goal [13,17]. Goal achievement also depends upon deliberate reflective thought about how to improve a skill [18–22].

This study focused on how to improve self-efficacy and questioning skills of preservice teachers by implementing best practices in feedback and coaching. The context was an undergraduate course enhanced with a mixed-reality simulation (MRS) in which candidates were expected to develop and present HOT questions in each of three 10 min lessons. We hypothesized that teacher candidates would improve self-efficacy for using questioning skills when data-driven feedback about the numbers and types of questions asked during a lesson was followed by coaching that included tailored guidance to improve these skills.

2. Lesson Delivery in a Mixed-Reality Simulation Environment

Opportunities to observe teacher candidates rehearse high-leverage practices typically occur in actual PreK-12 classrooms or simulated environments. Numerous subject area requirements and expanding regulations for visitors in most public schools, however, limit the time practicing teachers can spend in actual classrooms. Many administrators of teacher preparation programs have had to enact other options prior to student teaching. Fortunately, an MRS environment allows a preservice candidate to teach a lesson within a virtual classroom [23,24].

Empirical studies in virtual environments for preservice teacher education are gaining recognition [25–31] Combining this innovative technology with best practices in teaching will add to our knowledge about effective strategies for educating preservice teachers.

The mixed-reality simulation system used in the current study was Mursion®. This platform allows a preservice candidate to teach within a simulated classroom environment. Whether teaching online or in a university classroom, the monitor displays a virtual class complete with simulated students (Figure 1). The camera is used to track where the candidate is located with respect to the monitor, the room, and the students in the virtual world. This gives the impression of teaching in a realistic environment. To capture the teaching experience, all lessons are recorded, generating an archive of a participant’s performance.



Figure 1. Snapshot of recorded video of simulated students that the preservice teacher sees. *Note.* Copyright 2023 Mursion, Inc. Reproduced with permission from Mursion®. *Middle-school simulated students* [32].

The mixed-reality classroom contains five simulated students, each with their own ethnicity, personality, behaviors, and learning needs [23,24]. On request, the parent company of the simulation can vary the students to reflect different ages, behavioral scenarios, and content areas, tailoring the experience for each preservice teacher. The classroom students are people trained to respond in real time and are referred to as avatars. In this study, the classroom scenario using the high-order questioning skills was developed by Mursion® [32].

3. Related Literature

3.1. Self-Efficacy and Reflection

Learning to teach in a real or simulated classroom involves a complex process that integrates targeted performances, feedback about those performances, and self-reflection, which are the foundational components of [33] social cognitive theory of self-efficacy. This theory includes four dimensions for how an individual receives information about a performance: the mastery of target skills, observational learning, the amount and type of specific encouragement received, and the presence of specific physiological and emotional states, such as anxiety [14–16]. When individuals encounter any of these dimensions, self-reflection is used to adjust a future performance to obtain the desired goal. Bandura [18,33,34] explained that, as people gain insights into their own performances through self-examination, they become agents for their future actions. Strategies to improve reflection or metacognition are especially needed for preservice teachers [35]. Successful approaches include reflective journals and discussions [21,36].

The act of reflection is embedded in the teaching–learning process through questions and dialogue that solicit thoughtful responses. Dewey [37] was the first to emphasize the value of reflection for teachers. Schön [38] described types of reflective practices that have been systematically researched (e.g., Boody [19]; Killion and Todnem [20]; Liu [21]). Reflection typically includes thinking about the knowledge needed to enact a task, reflecting “on your feet” during an activity, and reviewing what occurred [22]. Boody [19] believes reflection is a powerful tool that depends on our level of known information and the reconciliation of that information with evidence. In Boody’s model, reflection falls into one of four categories: retrospection, problem solving, critical reflection, and reflection-in-action. Reflecting as retrospection requires reconsidering and learning from prior experiences. Problem solving allows an individual to place oneself in a prior event to think about how a situation might have been handled differently. When a similar experience happens again individuals can adjust accordingly. Critical reflection, according to Boody, is “exploring what is most educationally worthwhile and creating the conditions that would allow all people to equally join in the dialogue on what is of most worth” [19]. This style of reflection focuses on the educator’s actions within the immediate education setting. Whereas reflection-in-action occurs during a current event or situation, in a time where one may change the outcome in a meaningful way. In-the-moment reflection can be challenging for new teachers due to lack of experience. Boody’s framework was selected for this study because these four categories represent critical areas of reflection that could easily be included into a coaching schema for preservice teachers.

3.2. Inquiry Begins with Asking Questions

“Inquiry is learning by questioning . . . ” [3]. When teachers ask questions that provide opportunities for students to reflect on their responses and engage in dialogue about ideas, they are modeling inquiry behaviors [39]. Although creating a student-centered classroom based on inquiry teaching strategies can be a complex process [40,41], asking HOT questions is a direct path to entering the frame of mind for inquiry learning in any content area [42].

Providing data-driven feedback and coaching for classroom teachers in live settings improves the number of HOT questions asked by both the teachers and students [43]. When teachers intentionally allotted the time to insert HOT questions into their lesson

plans, they reported greater ease incorporating HOT questions into classes, and lesson delivery became more rewarding because students actively engaged in responding to course content. Using a mixed-reality simulation, Dieker et al. [44] had similar success training middle-school mathematics teachers in high-leverage instructional strategies when they checked for transfer from the simulated environment to actual classroom performance.

3.3. Feedback and Coaching in Education Using Non-MRS and MRS Environments

When developing our research design for supporting preservice teachers, we wanted to understand the best practices for using feedback within a coaching model. Hattie and Timperley defined feedback as “information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one’s performance or understanding” [10] and described a model of feedback that includes: (a) specific performance data about a task (task level), (b) information regarding the process for obtaining a satisfactory performance (process level), (c) explanations about how to monitor one’s performance (self-regulation level), and (d) personal appraisals directed toward the individual performing the task (self level). To effectively promote change, these four elements should be systematically employed in a coaching model. Despite a wide variety of coaching paradigms [45], Joyce and Showers were among the first to characterize coaching as “an observation and feedback cycle in an ongoing instructional or clinical situation” [46]. Kraft et al. [45] maintained that, in an optimal coaching model, sessions should be: (a) individualized, not group meetings, (b) scheduled with meaningful intervals to provide intensive training, (c) sustained over at least a semester, (d) context-specific to a teacher’s immediate environment, and (e) focused on specific skills. These components provide opportunities for dialogue and reflection. Kraft and Blazar [13] successfully implemented these coaching characteristics to improve classroom management and instructional practices. Coached teachers were rated significantly higher than control teachers on the ability to challenge students with rigorous work. Although the effects of direct feedback through coaching appear to be promising, Kraft and Blazar cautiously concluded that these results might not apply to a larger context or another coaching model.

When Averill et al. [17] examined the impact of rehearsal teaching and used in-the-moment coaching with preservice candidates to encourage high-leverage practices of mathematical thinking and communication, candidates valued both the prompt response about the lesson and the ability to immediately implement the coach’s suggestions. Khalil et al. [47] also used feedback and implemented rehearsal teaching with prospective mathematics teachers. Using an MRS environment, these researchers successfully implemented a cycle of teaching, feedback, and re-teaching practices that enabled preservice teachers to enact instruction in a controlled and monitored setting, allowing them to improve their instruction. The flexibility of rehearsal teaching in the virtual classroom was especially advantageous. Multiple training trials were also implemented by Garland et al. [48] who successfully instructed a small group of special education teachers. Multiple virtual reality simulation sessions, including expert coaching with feedback and demonstration, led to improved performance and confidence by employing a specialized classroom strategy. Pas et al. [49] used avatars to adjust to requests by special educators for coaching with specific feedback in classroom management strategies. These researchers reported improvement in teachers’ skills and positive ratings by both participants and coaches about the simulated environment. Each of these studies had successful results that provided guidance on how to conceptualize feedback as a distinct entity in a coaching model, although some studies provided more tangible descriptions of feedback [44,48] than others [17,47]. Only one other study focused on providing an intervention for preservice teachers implementing HOT questioning techniques in an MRS environment [29]. While Mendelson and Piro provided various forms of feedback in the intervention, they did not provide direct coaching as used in the context of the current study.

Believing that feedback should be as closely related to the task as possible, we decided to address implementing a high-leverage inquiry practice through the discrete activity

of asking questions and by collecting data to inform preservice teachers about the types and numbers of questions they asked while presenting each lesson. We also provided systematic coaching to improve the use of HOT questions for these teacher candidates.

4. Research Questions

We used the term data-driven feedback to express that the content of the feedback was based on observed quantifiable data, rather than more subjective information. Our coaching schema was created to improve HOT questioning skills and teacher self-efficacy over three 10 min teaching sessions in a preservice teacher education course enhanced by an MRS environment. To assess effectiveness, we asked: (a) Is there a difference in preservice teachers' self-efficacy between those who received data-driven feedback and coaching to improve questioning skills and those who did not? (b) Is there a difference in the number of K/C and HOT questions asked by preservice teachers in the treatment and the comparison groups? (c) For preservice teachers in the treatment and comparison groups, what are their perceptions of incorporating HOT questions into their lessons?

5. Materials and Methods

5.1. Research Design and Procedures

Each of the two course sections enrolled 15 preservice candidates. Sections were randomly assigned to either a treatment or a comparison group [50]. This research followed a mixed-methods exploratory sequential design in which a quasi-experimental process was used to examine research questions one and two. A case-study design was employed to address research question three [51]. For the first research question, participants completed the Teachers' Sense of Efficacy Scale (TSES) [52] at the beginning and end of the semester. Means for the two groups were compared.

For research question two, baseline data were obtained from the prior semester regarding candidates' use of questions when teaching in the mixed-reality simulator. Paced throughout the following semester, participants in the treatment and the comparison groups were expected to prepare and deliver three 10 min lessons. A high-leverage practice was assigned to all students in each course section: include a HOT question in each lesson. All candidates received the same written explanation of the high-leverage practice and examples of HOT questions. Candidates developed their own lessons related to their area of certification, that is, elementary education, secondary chemistry, etc. The lessons occurred in a classroom containing the MRS technology and "fish-bowl" seating in which classmates, a course facilitator, and a researcher viewed every session. The role of the course facilitator was to (a) ensure that the candidates observed the time schedule, and (b) provide a brief verbal summary about the performance of each preservice teacher in both groups. A researcher attended each session to monitor the simulation hardware, operate the videorecording equipment, and take notes.

Other students occasionally provided remarks to a candidate after a lesson. In summary, members of both groups received written documentation about composing HOT questions; created and taught three lessons using the virtual reality simulator; when time allowed, received a brief verbal summary of their lesson performance from a course facilitator; received data-driven feedback about K/C and HOT questions from each session (treatment—during coaching; and comparison—at the end of the study); and participated in a post-study interview.

Only the treatment group members were contacted prior to the first lesson and after each lesson presentation. Before teaching the first lesson, an initial conversation included introductions, and a discussion about the different types of questions, K/C and HOT. One to three days after a lesson, a researcher contacted each participant to ask how they thought they performed in lesson delivery and convey the data-driven feedback of the numbers and types of questions that the participant asked in the recent lesson. Coaching then ensued as the conversation shifted to recognizing differences between types of questions, asking the candidate to compose a HOT question for the next lesson, and discussing strategies to insert

the HOT question into the subsequent lesson. These conversations were audio recorded. K/C and HOT questions from each lesson taught within the MRS were compared by group to address the second research question. Coaching sessions included the following protocol:

Contact was made with the participants in the treatment group through phone calls or Internet conferencing systems, such as Zoom or FaceTime. It was planned that each session should not exceed 30 min. End-of-semester interview data were used to gain insights into the perceptions of all treatment and comparison group students. The cases were bound by assignment to each group. The coded data from all data-driven feedback/coaching sessions and end-of-semester interviews were used to address the third research question.

5.2. Research Context and Sample

This study was conducted at a medium-sized state university in northeastern United States. Over 85% of the student body were in-state residents, and 145 were enrolled in a nationally approved initial certification program in education [53]. All 30 undergraduates completing their third or fourth course in the education sequence consented to participate. The same professor taught both sections of the course; activities related to the mixed-reality simulation were an ungraded component of the syllabus. The grade levels candidates were preparing to teach ranged from elementary to secondary.

5.3. Instrumentation and Data Collection Tools

Self-efficacy was assessed using the Teachers' Sense of Efficacy Scale (TSES) [52], a 24-item survey with three subscales: efficacy in student engagement, instructional strategies, and classroom management. Responses use a 9-point Likert scale to record beliefs about one's current ability to perform a specific task. Tschannen-Moran and Hoy suggested that scores below 3 be considered low, and scores above 6 as high. When interpreting the TSES with preservice teachers, they recommended using total mean scores, not individual subscales, because preservice teachers' responses are often less distinct among the subscales compared to those of experienced teachers. Overall internal consistency reliability was reported at 0.90 [54]. "The high correlations between teacher efficacy beliefs and teaching strategies indicated that the TSES had good predict[ive] validity" [54]. At the beginning and end of the semester, both treatment and comparison groups completed the survey in 10 min.

Demographic data were collected via questionnaire. In the treatment group, 1 participant identified as male, 14 as female. In the comparison group, 5 candidates identified as male and 10 as female. Descriptive data regarding ages and GPAs are in Table 1.

Table 1. Student demographic mean data.

Participants	Age	GPA
Treatment		
Elementary (K-6)	20.2	3.6
Secondary (7–12) Chemistry, Health, Mathematics, Spanish	21.6	3.4
Total	21.0	3.5
Comparison		
Elementary (K-6)	24.7	3.5
Secondary (7–12), Chemistry, Health, Music	20.3	3.7
Total	22.5	3.6

NOTE: The treatment group started with 17 members; 2 were removed due to incomplete data. GPAs are on a 4-point scale.

All sessions were observed live by a researcher, videorecorded, and notes were taken for all verbal exchanges. Verbal interactions from the videos were transcribed by a professional typist. By using both the transcript and the visual image when coding, we could

precisely measure the wait-time when questions were employed by the teacher candidates. These data were corroborated by another researcher.

In addition to videorecording each session, the Classroom Practices Record (CPR) [55] was used to register all verbal interactions between the preservice teachers and students. These exchanges were used to record data about the frequency of K/C (knowledge/comprehension) and HOT questions (i.e., Bloom's application, analysis, synthesis, and evaluation levels of thinking) generated by a teacher candidate in each lesson. To maintain coding accuracy, "The observer does not distinguish among these higher levels; rather, the 'HOTS' acronym [is used] to note that a higher order thinking skill question was raised" [55]. While the CPR can be used to record and code both questions and responses, questioning skills were the focus of this study. We chose to limit the expectations for the candidates, given the short time frame of 10 min for a single lesson.

The original CPR protocol underwent rigorous standardization measures including a process to establish observer coding agreement that "met or exceeded the 0.80 criterion" [55]. The same dichotomous coding procedures for identifying levels of questioning skills were employed in the current study and generated 100% agreement between two of the researchers. Labeling each HOT level based on the taxonomy is an optional activity.

The feedback and coaching protocol was field-tested at two different sites. One to three days after a lesson presentation, a researcher contacted each participant in the treatment group via phone or Internet conference service to present the data-driven feedback from the CPR, discuss lesson performance, and plan questioning strategies for the next teaching session. The following steps were followed in each coaching session:

1. (Initial coaching session) Refer to the demographic form, I see that you are in the elementary education program/secondary education program. What topics do you look forward to teaching when you complete the program?
2. How do you think the MRS experience is preparing you for student teaching?
3. In your most recent MRS session, how do you think you did?
4. How do you think you did with respect to the questions you asked?
5. What seemed easy to do? What seemed difficult?
6. At this time, the researcher explained the number of K/C and HOT level questions the participant generated during the lesson. After the results were given, the researcher answered any questions to make sure the participant understood the results.
7. What do you think about these results? How were you thinking about the types of questions you were asking during the session?
8. What do you think you could do to improve your questioning skills? What do you think you could do to include more HOTs questions in your next session?
9. Do you need any additional resources to help you achieve your goal?
10. Recommendations for next session: At this time the researcher gave some strategies for improving their questioning skills and explained how HOT questions can be used to promote student engagement. Generic examples of levels of questions were used to help candidates differentiate between K/C and HOT questions. The researcher did not review questions for the next lesson. Students initiated their own questions when they individually planned their lesson activities.
11. Finally, the call ended with closing rapport.

There were 45 separate 15 to 25 min data-driven feedback and coaching sessions with the treatment group members. Each call was audio recorded and then transcribed.

After the three sessions were completed and the TSES was administered, each participant was interviewed via phone. Interviews were digitally recorded. Interview questions 1 to 5 prompted reflections on using questioning strategies over the semester and lesson preparation techniques, perspectives of in-class performance, insights about the student and preservice teacher interactions, and changes they made in their teaching. Questions 6 through 8 addressed the coaching they received, including thoughts about the process and how it affected their planning and overall performance. Questions 9 and 10

asked what advice they would give to the next cohort of students and how the mixed-reality simulation experience could be improved. Each interview lasted 20 to 25 min.

6. Analyses

Research question #1 was addressed using ANCOVA to compare group differences for TSES mean scores. For research question #2, we had planned an ANOVA to analyze the mean differences between conditions, however, the low number of HOT questions from the comparison group meant that the data were no longer at the interval level. Therefore, a two-sample case chi-square test for independent samples was conducted in which the frequency for the type of question asked, HOT and K/C, and the type of research condition data, driven feedback and coaching versus no data-driven feedback and coaching, served as the variables [56]. First- and second-cycle coding schemes recommended by [57] were used to analyze interview responses for research question #3.

7. Results

7.1. Research Question #1

Prior to assessing self-efficacy between groups, academic equivalency was established by comparing mean Grade Point Average (GPA) values. Both groups had mean GPA values of B+ and there was no statistically significant difference between GPAs for the treatment ($M = 3.49$, $SD = 0.23$) and comparison groups ($M = 3.65$, $SD = 0.25$); $F(1,27) = 3.22$, $p = 0.08$, and $\eta^2 = 0.10$. Therefore, achievement levels between groups were not a consideration affecting the findings.

Most participants were completing their third course using the mixed-reality simulator. When the TSES [52] was administered at the start of the semester, the average score for each group was high, above six on a 9-point scale (Table 2). After assuring that all the assumptions were met, an ANOVA between the treatment and comparison groups' TSES pretest scores indicated no statistically significant difference between conditions; $F(1,28) = 2.94$, $p = 0.10$, and $\eta^2 = 0.10$. The groups were equivalent with respect to teacher self-efficacy at the start of the study. By the end of the 15-week semester, TSES mean scores for both groups improved slightly. After analyzing the assumptions for the posttest scores, the test for the equality of variances was statistically significant, $p = 0.04$. To address the outcome of unequal variances, we conducted an ANCOVA, as recommended by Hinkle et al. [56], using pretest scores to adjust the posttest scores for error. The ANCOVA revealed no statistically significant difference between the groups; $F(1,25) = 0.66$, $p = 0.42$, and $\eta^2 = 0.03$. A question on the TSES that specifically addresses the focus of this study is "To what extent can you craft good questions for your students?" There was no statistically significant difference between the groups for their mean responses to this item on the pretest; $t(28)$, 1.262, and $p < 0.15$ (treatment mean = 7.06, and comparison mean = 6.47), but there was a statistically significant difference between means; $t(28)$, 1.545, and $p < 0.07$, for the posttest, with the treatment group mean ($M = 7.41$) being significantly higher than the comparison group mean ($M = 6.46$).

Table 2. TSES Scores.

Group	Treatment			Comparison		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Pretest	15	7.39	0.93	15	6.76	1.07
Posttest	15	7.47	0.55	13 *	6.98	1.01

NOTE: * Two participants in the comparison group did not complete their simulations within the course time frame.

7.2. Research Question #2

In the semester prior to this study, members of both groups used the MRS in their preservice class. The high-leverage expectation was to teach a lesson about using a graphic organizer. Videos were available for 12/15 candidates from the treatment group and 7/15 from the comparison group for one session. Members of the treatment group asked a mean of three K/C questions (range, 0–6) and no HOT questions. The mean number of K/C questions asked by those in the comparison group was two (range, 0–6) with only one HOT question recorded. The *t*-test revealed that there was no significant difference between the number of K/C questions asked, prior to the commencement of the study ($t = 1.41$, and $p = 0.18$).

By the end of this study, a total of 112 K/C and 54 HOT questions were asked by the treatment group members and 148 K/C and 4 HOT questions were recorded for those in the comparison group. Sample K/C questions included: Where do plants get their food? How long do you hold a quarter note? How many systems are in the human body? In comparison, a sample of HOT questions are represented by the following: What would happen [to the eco-system] if sharks became extinct? If you only ate one color of food for a month what [do you think] would be the best food to eat? How does the amount of precipitation in an eco-system effect what lives there? The chi-square analysis revealed a statistically significant difference between K/C and HOT question performance for the treatment and comparison groups; $\chi^2(47.46)$, $p < 0.01$, and Cramer's $V = 1.26$. Additionally, all four residuals were important contributors to the chi-square value, ranging from an absolute value of 2.04 to 4.51. The total of 54 HOT questions asked by those in the treatment group, was well above the expected value of 30.28 (residual = 4.31). The number of K/C questions for this group was 112, falling below the expected value of 135.72 (residual = -2.04). However, the number of HOT questions observed for the comparison group was four, considerably below the expected amount of 27.72 (residual = -4.51), generating the largest residual value in the chi-square procedure. Comparison group members asked more K/C questions than expected (observed = 148, expected = 124.28, and residual = 2.13).

Throughout this study, when observing the ratio of question types per session, members of the treatment group shifted toward using more HOT questions in sessions two and three, demonstrating that they were more likely to engage in best practices for posing questions than their peers in the comparison group. This incremental increase for members of the treatment group is also observed in the number of HOT questions asked and the number of candidates engaged in posing these questions from sessions one to three. Ten candidates in session one asked 17 HOT questions, however, five participants asked only K/C questions. In session two, 18 HOT questions were asked by 12 candidates in the treatment group, with no HOT questions asked by three of the candidates. In session three, 19 HOT questions were asked by 15 preservice teachers with each asking between one to three HOT questions. In session one, the ratio between K/C and HOT questions was 3.2 K/C per 1.13 HOT. This ratio changed to 1.80 K/C per 1.20 HOT for session two and 2.47 K/C per 1.27 HOT, for session three. The 10 min limit of the session and the increased time it takes to use HOT questioning practices restricted how much this ratio could change from session two to three. The HOT questions took longer to execute, when performed correctly, which afforded less time for the candidate to ask any other questions. On average, the ratio for a 10 min session appears to be 2.0 K/C to 1.0 HOT. In contrast, the comparison group generated mostly K/C questions and did not incrementally increase their use of HOT questions. The comparison group's ratio between K/C and HOT questions for session one was 3.33 K/C per 0.07 HOT; session two resulted in a ratio of 3.40 K/C per 0.13 HOT; and for session three the ratio was 3.36 K/C per 0.07 HOT. Without insights provided by coaching, comparison group members did not have the support to improve their HOT-question performance, suggesting that the initial written explanation of HOT question formation did not help to develop this skill.

After a review of the video recordings revealed differences between the delivery of the HOT questions across participants, we created a categorization system of three levels using the following components: (a) higher-order thinking, (b) wait-time, and (c) engagement. A Level 1 HOT question met the minimum requirement of representing higher-order thinking. A sample HOT science question was “What would life be like if the world was still like Pangea?” However, asking the question and immediately continuing with the lesson did not allow classroom students the opportunity to think about a possible response. Therefore, a Level 2 HOT question entailed providing at least three to five seconds of wait-time in addition to achieving Level 1 status. Although some participants posed a HOT question (Level 1) and waited for students to think about how to respond (Level 2), they did not use the response to elicit dialogue and engagement amongst the class members in the MRS. If at least one student responded and another student made a meaningful contribution to the conversation, the question was designated as Level 3.

The treatment group’s questions in the first session were distributed between HOT question Levels 1 and 2, with the majority being Level 1. In the second and third sessions, the treatment group’s questions were distributed between HOT question Levels 1, 2 and 3. Members of the treatment group shifted toward using the higher leveled HOT questions in the final two sessions. The shift in types and levels of questions for the treatment and comparison groups is indicated in Tables 3 and 4. However, the overall ratio of K/C and HOT questions changed little between these two sessions, remaining at approximately 2.0 K/C to 1.0 HOT. After weighting the HOT questions (Level 1 = 1; Level 2 = 2, and Level 3 = 3), a matched-pair sign test was used with an a priori significance level of $p < 0.10$ for exploratory purposes. The only statistically significant change in the average number of HOT questions was from session one to session two for the treatment group, $p = 0.08$.

Table 3. Treatment group question performance for each session.

Participant	Session One Questions					Session Two Questions					Session Three Questions				
	K/C	HOT Level 1	HOT Level 2	HOT Level 3	Total HOT	K/C	HOT Level 1	HOT Level 2	HOT Level 3	Total HOT	K/C	HOT Level 1	HOT Level 2	HOT Level 3	Total HOT
T01	4	0	1	0	1	0	0	1	0	1	1	2	1	0	3
T02	4	0	2	0	2	3	1	1	0	2	3	1		0	1
T03	3	2	0	0	2	1	0	0	1	1	2	0	1	0	1
T04	3	1	1	0	2	0	0	1	1	2	1	1	0	0	1
T05	4	3	0	0	3	0	0	2	0	2	1	0	1	0	1
T06	2	0	1	0	1	2	0	1	0	1	0	0	0	1	1
T07	3	0	1	0	1	2	1	0	0	1	2	0	1	0	1
T08	5	0	0	0	0	2	0	1	0	1	3	1	0	0	1
T09	2	0	0	0	0	3	0	0	0	0	3	1	0	0	1
T10	2	1	1	0	2	0	2	1	0	3	3	2	0	0	2
T11	4	0	0	0	0	2	1	1	0	2	4	2	0	0	2
T12	3	2	0	0	2	4	0	0	0	0	4	0	1	0	1
T13	2	0	0	0	0	5	0	0	0	0	3	1	0	0	1
T14	4	1	0	0	1	1	0	1	0	1	2	0	1	0	1
T15	3	0	0	0	0	2	1	0	0	1	5	0	0	1	1
Total	48	10	7	0	17	27	6	10	2	18	37	11	6	2	19
Average	3.20	1.67	1.17	0	1.13	1.80	1.20	1.11	1.00	1.20	2.47	1.38	1.00	1.00	1.27
Ratio	3.20 K/C: 1.13 HOT					1.80 K/C: 1.20 HOT					2.47 K/C: 1.27 HOT				

Table 4. Comparison group question performance for each session.

Participant	Session One Questions					Session Two Questions					Session Three Questions				
	K/C	HOT Level 1	HOT Level 2	HOT Level 3	Total HOT	K/C	HOT Level 1	HOT Level 2	HOT Level 3	Total HOT	K/C	HOT Level 1	HOT Level 2	HOT Level 3	Total HOT
C01	5	0	0	0	0	4	0	0	0	0	3	0	0	0	0
C02	6	0	0	0	0	5	0	0	0	0	5	0	0	0	0
C03	2	0	0	0	0	3	0	0	0	0	4	0	0	0	0
C04	1	0	0	0	0	4	0	0	0	0	3	0	0	0	0
C05	5	0	0	0	0	2	1	0	0	1	4	0	0	0	0
C06	3	0	0	0	0	2	0	0	0	0	2	0	0	0	0
C07	4	0	0	0	0	5	1	0	0	1	4	0	0	0	0
C08	2	0	0	0	0	2	0	0	0	0	1	0	1	0	1
C09	4	0	0	0	0	4	0	0	0	0	NA	NA	NA	NA	NA
C10	1	0	1	0	1	4	0	0	0	0	5	0	0	0	0
C11	4	0	0	0	0	3	0	0	0	0	3	0	0	0	0
C12	5	0	0	0	0	4	0	0	0	0	4	0	0	0	0
C13	4	0	0	0	0	3	0	0	0	0	4	0	0	0	0
C14	3	0	0	0	0	3	0	0	0	0	3	0	0	0	0
C15	1	0	0	0	0	3	0	0	0	0	2	0	0	0	0
Total	50	0	1	0	1	51	2	0	0	2	47	0	1	0	1
Average	3.33	0	0.13	0	0.07	3.40	1	0	0	0.13	3.36	0	1	0	0.07
Ratio	3.33 K/C: 0.07 HOT					3.40 K/C: 0.13 HOT					3.36 K/C: 0.07 HOT				

Although the CPR observation tool was validated on the use of a dichotomous coding schema for labeling types of questions, K/C or HOT, we identified the specific level of Bloom's taxonomy for each HOT question asked. Most questions for the treatment group were at the analysis level, 34/54 questions. Three were categorized as application, nine as synthesis, and eight as evaluation. For the comparison group, one was identified as application, two represented analysis, and one reflected evaluation. Further analysis revealed that the HOT questions categorized at the evaluation level, the top tier in the taxonomy, included four Level 1 and four Level 2 HOT questions, as well as one HOT question at Level 3. The only evaluation question for the comparison group was at Level 2.

7.3. Research Question #3

Based on the responses to the interview questions, 24 codes and 13 categories were developed which were categorized into four main themes, described below. Theme 1 pertains to the treatment group only.

Theme 1 was labeled data-driven feedback and coaching improve perceptions of self-efficacy for creating and implementing HOT questions. As noted in the results for research question one, there were no statistically significant differences in the overall means on the TSES [52] between groups, however, the treatment group had a higher posttest mean for the one item related to "crafting good questions". As noted in the analyses for research question two, prior to the study, there was no statistically significant difference in K/C question formation between groups and only one HOT question was posed, yet at the completion of data collection, the treatment group's performance in creating and applying HOT questions was overwhelmingly better than that of the comparison group. In fact, every member of the treatment group incorporated one to three HOT questions into at least one of the three lessons, indicated improvement in HOT question creation, acknowledged the amount of social encouragement received through the coaching process, and reported feeling more confident in implementing HOT questions into lessons by the end of the third session. Categories of responses during coaching sessions with the treatment group as related to self-efficacy are reported in Table 5.

Table 5. Frequency of codes from the treatment group’s coaching sessions.

Categories Related to Self-Efficacy/Codes	Session 1	Session 2	Session 3	Total
Mastery of experiences				
Connects questioning skills to developing good teaching practices	5	17	19	41
Develops clearer questioning goals after coaching session	2	9	15	26
Believes the simulation is a positive experience for developing questioning skills	6	8	3	17
Amount or type of social encouragement received				
Acknowledges coaching had a positive effect on lesson implementation	0	23	21	44
Appreciates one-on-one coaching	0	15	15	30
The individual’s responses to stressors				
Expresses confidence, pride, and relief in “doing a lesson”	0	10	11	21
Displays low confidence in performing the lesson	3	1	1	5

NOTE: There were 15 participants in the treatment group (data-driven feedback and coaching).

Treatment participant 12 (T12) demonstrated how questioning strategies provided insights into becoming a better teacher when he stated, “I feel like you have to ask them a [HOT] question. At least that will get them to critically think about what you’re asking them . . . to get the kids to talk”. T14 appreciated the social encouragement from coaching when she stated, “It’s really making it a lot easier having someone to talk to about it . . . and to be able to discuss it afterwards to see where we can see what needs to improve”. Realizing that students were responding within the lesson boosted T06’s confidence; she said, “My higher-level thinking question [about] whether exercise or [healthy] eating had a greater impact on your body . . . I felt the [avatars] responded to that”. Instead of just delivering a lesson, treatment participants recognized their progressive mastery of creating HOT questions for their lessons and overcame their hesitancy to ask questions and engage in conversations.

Theme 2 was identified as planning for a lesson requires reflection. Treatment group members credited the coaching sessions for helping them develop better planning skills. From the first individual meetings, before a researcher had performance data to share, treatment group members asked questions about how to create and use HOT questions in their lessons. After each session in the MRS, these candidates reported using the data-driven feedback and coaching to reflect on how to improve questioning skills. At the end of the semester, these participants indicated that they increased their lesson planning time after each coaching session ($n = 29$ responses) and composed better questions on their own between sessions ($n = 22$ responses). The participants also remarked about how they thought a question worked in a former lesson and how to write a better question, indicating the use of reflection as retrospection and problem solving. For example, some treatment group members made the following comments:

I drew different ideas and different questions from . . . different lessons to kind of make up my own [questions] and then I also developed the answers to the questions myself, just in case. (T03)

I was trying to find my own ways to connect [my questions] with the students. (T05)

[Two of the avatars] really like reading and I just take . . . the individual’s personality and try to find a way to incorporate it all in one lesson so that they’re all engaged. (T15)

Unlike the treatment group, the researcher contacted the comparison group only at the end of study, during the phone interview. At this time, all comparison group members received data-driven feedback about their performance using questioning skills throughout the semester. When asked about their planning procedures, 12 comparison group members stated they did not change their planning methods based on their simulator experiences, which is illustrated by this statement from comparison group member C11, “[For each

lesson], I would write a little bit of a script . . . usually an introduction that I would stick with almost word for word”.

Theme 3 was best depicted as lesson performance is enhanced by reflection. All members of the treatment group responded that lesson performance improved over the semester. When performing their lessons, they engaged in predominantly two types of reflection about their performance, reflection-in-action and critical reflection. Across all interviews, treatment group members commented 29 times that they were able to adapt to unplanned interactions (reflection-in-action) in the lessons with the avatars, while those in the comparison group indicated 19 instances of classroom management issues and problems adapting to unexpected student interactions. Treatment group members critically reflected that the simulated students responded well to their lessons with very minor issues in classroom management ($n = 19$) and explained how their performance improved over the semester ($n = 30$). T08 commented, “I feel like they didn’t fool around as much because they were thinking about the questions”. T07 said, “I feel [that] the classroom management, lesson, and . . . my organization of the lesson [have] gotten better”. Treatment group members entered a lesson feeling well prepared with a clear plan for lesson objectives and strategies to engage student learning through the use of questions.

Most comparison group members ($n = 10$) stated that their overall lesson performance did not improve very much over the semester, as indicated by this comment: “I don’t think it’s changed a ton. I think maybe like classroom management got a little better, maybe that’s better” (C06). Despite this having been their third semester using the mixed-reality simulation, some members of the comparison group noted that it was difficult to use the designated lesson time. Comparison group member C09 said, “I’ve left a couple of the [mixed-reality] lessons thinking, ‘I ran out of time or I didn’t reach all the points I wanted to,’ and there were times where I wish I had a little more time, but also times where I wish I had a little less time”. Without direction from a coach, or another type of intervention, their ability to think critically about their performance or engage in reflection-in-action appears to have been limited.

Theme 4 was represented by the statement that data-driven feedback and coaching improves questioning skills. Every member of the treatment group recognized personal growth in lesson development techniques through engagement with HOT questioning strategies and described how these strategies improved over time ($n = 27$ responses) due to their participation in the study. Their comments included:

Well, I feel like I’ve explored more with the whole higher-order thinking questions. I didn’t really ever put much thought into that before this semester. (T01)

Going from asking questions where I had the specific answer in my head to asking questions to see what they would think . . . it was smoother in being able to communicate the questioning and have the kids as engaged as possible. (T13)

I think it was effective and I think that was shown in my assessment [from the coach’s feedback] on the last session when I reflected on the first session, all the kids were able to tell me what they did and their answers came a long way from the first session. (T03)

As a result of the coaching, these candidates’ ability to label questions as K/C or HOT became routine. Two students recognized that, after they posed a HOT question, the avatars created their own HOT questions in response. Although the purpose of this study was to monitor changes in the questioning skills of preservice teachers, a logical extension of this methodology is to record K/C and HOT questions and responses for both teachers and students [43,55].

Although members of the comparison group were given initial written material about developing HOT questions, they felt their overall questioning skills were unchanged ($n = 9$). Some even stated they felt unclear about what a true HOT question was ($n = 13$), or where to use one in a lesson. In their comments, they barely mentioned growth in question creation through statements, such as the following:

I didn't always get to them [the questions] because of the way I structured my lesson for the amount of time we had. (C15)

Higher-order thinking, well again that's a little more difficult with music because we really don't use higher-order thinking questions. (C14)

The way I did it, I taught the content first and then I kept asking questions to see if they understood the content. (C06)

It's hard to plan for higher-order thinking . . . it can only come naturally. (C10)

8. Discussion

Self-efficacy is the level of certainty individuals have about their ability to perform an action or make decisions to achieve desired results [15]. The treatment in this study was based on Bandura's [14–16] notions to improve skill mastery, thereby enhancing self-efficacy, by focusing on (a) the high-leverage practice of employing questioning skills during lessons over a 15-week semester; (b) peer observation; (c) social encouragement received from the researcher who provided data-driven feedback to highlight the numbers and types of questions incorporated into each lesson; and (d) a decrease in stressors for the preservice candidate provided during each supportive coaching session. To actualize Bandura's theory, candidates in the treatment group should become more self-efficacious by improving task performance over time, learning from peers, receiving encouragement, and experiencing reduced stress when performing the target task. By the end of this study, the treatment group did show higher confidence to "craft good questions" for their students (research question 1) and were more competent in delivering high-quality HOT questions (research question 2) as compared to the comparison group. Regarding peer observations, although students watched their classmates teach in front of the classroom of avatars, these lessons were not mentioned by any participant as a contribution to enhance candidate learning. Treatment group members felt reinforced by the coaching, while, for comparison group members, the summary from the facilitator at the end of each lesson did not appear to influence their performance in HOT question development. Finally, responses to the final interviews indicated that the members of the treatment group felt more accomplished over time when teaching a lesson, feeling that the avatars were responding to their questions and engaging in the content, and exhibited signs of reduced anxiety in lesson delivery. Unfortunately, their comparison group peers mentioned that their teaching formats had not changed, with many feeling uncertain about the use of HOT questions.

8.1. Research Question #1: Unchanged Mean Teacher Self-Efficacy Scores between Groups

Despite having a treatment designed to address Bandura's four dimensions of self-efficacy [14], there was no statistically significant difference in general teaching self-efficacy between the treatment and comparison groups as determined by the average posttest scores on the TSES. Potential explanations for this nonsignificant result include possible ceiling effects for both groups, a false sense of success on the part of the comparison group members, and the need for greater statistical power to detect group differences in the target skill. A ceiling effect could be explained by the fact that both groups began the semester with high pretest scores in self-efficacy regarding their general capabilities as teacher candidates. Pretest and posttest group means for both groups were above the high rating indicator of six on a 9-point scale. For most participants, this was the third teaching course of four, prior to being placed into student teaching, and it was also the third course in which the participants used the mixed-reality simulator, for a total of nine mixed-reality simulations. Candidates were used to interacting with the avatars, thus, based on past experience, they may have entered the semester with high self-efficacy toward teaching in a mixed-reality environment. Consequently, these students had reasons to be confident. Perhaps they rated themselves so highly at both the beginning and end of the semester because they could look back over the past three courses and see how far they had come in

their knowledge and skill for teaching. A better measure of their self-efficacy may have been to assess changes in the TSES over the entire teacher certification program [27].

Bandura explained that some outcomes could undermine growth in self-efficacy. “A resilient sense of efficacy requires experience in overcoming obstacles through perseverant effort” [15]. When treatment group members received data about the numbers and types of questions they used in each lesson, this feedback provided them with information not only about what they had achieved, but also about what they could improve. To overcome a lack of expertise in question formation, these participants received expert scaffolding to incorporate both K/C and HOT questions into their lessons. This feedback and coaching maintained their already high teaching self-efficacy. In contrast, comparison group members did not receive data about their performances until after completing the posttest. This absence of feedback may have given these candidates a false sense of success.

Although the TSES posttest scores were higher for the treatment group members, the low sample size likely limited the robustness of the analysis [56]. It should be noted that the treatment group members scored statistically significantly higher than their peers on one TSES item related to asking questions. Furthermore, the candidates’ qualitative responses reflected how components of self-efficacy impacted their progress in the specific task of implementing questioning techniques into their lessons.

8.2. Research Question #2: Performance Gap between Conditions

The results of research question two revealed that the treatment group’s ability to create and use HOT questions in their simulated lessons completely outpaced the comparison group as indicated by the statistically significant chi-square. Delivery of a lesson by a preservice teacher followed by appropriate feedback [10] and coaching can help improve the individual practices of an educator [13]. The feedback in this study was easy to tally and quickly understood by the candidates. Feedback about the specific task was then woven into the conversation between the coach and preservice teacher as they discussed how to design questions that the candidate subsequently devised and integrated into the next lesson. Planning questions and delivering them within the lesson are crucial because some teacher candidates plan questions but never actually use them [43], as corroborated by participant C15. Thus, data-driven feedback and coaching appear to have positively influenced the practices of teacher candidates to not only create HOT questions but ask them within a lesson.

Based on this process, the coach can be seen as the facilitator of meaning-making, which is explained by Laverick as a “process that moves a learner from one experience into the next with deeper understanding of its relationships with and connections to other experiences and ideas” [58]. Laverick’s remarks support the treatment group’s ability to use the feedback and coaching to create a deeper understanding of questioning skills which, in turn, resulted in more high-quality HOT questions.

8.3. Research Question #3: Perceptions of Accomplishments between Conditions

All members of the treatment group indicated that they improved in HOT question creation and proudly explained how those questions were inserted into the lessons across the three sessions. They acknowledged that coaching positively impacted their accomplishments and that they were able to connect the use of questioning skills to effective teaching practices. These statements clearly point to the fact that they improved their confidence to master the use of questioning skills as a direct result of the data-driven feedback and coaching [15], although this outcome was not specifically reflected in the overall TSES results. Total comments by treatment group candidates included 84 statements referring to mastery of skills, 74 regarding encouragement from the coach, and 21 indicating that the data-driven feedback and coaching sessions led to a reduction in general anxiety about teaching the lessons (Table 5).

Treatment group members credited the coaching sessions for helping them develop better planning skills. We propose that, because someone was keeping track of their

performance and actively engaging them in dialogue to improve their skills, they felt more inclined to devote more time to planning. They reported that they not only planned more, but planned better, because of the coaching intervention. This connects to reflection as retrospection and reflection as problem solving, as described by Boody [19]. Reflection as retrospection requires reconsidering and learning from prior experiences, and reflection as problem solving places an individual in a prior event to think about how a situation might have been handled differently. The coach helped to facilitate both types of reflection because treatment group members changed their habits and increased their lesson planning time. Participant examples included how they anticipated potential student responses and addressed student interests during lesson planning.

Twelve of the fourteen comparison group members explained that they did not change their planning methods after the simulator sessions. This could mean that they failed to engage in reflection, possibly because there was no specific coaching guidance [19], or they reflected, but did not have the skills to make appropriate changes. Unlike the treatment group, without someone giving them guidance and monitoring their progress, the comparison group members might not have treated the simulation as seriously, and therefore may not have engaged in useful reflective practices to improve their skills. In contrast, the coach for the treatment group members was able to impel those participants to improve their questioning skills, despite the fact that there was no grade recognition for completing a lesson in the simulated classroom.

All treatment group members responded that their lesson performance improved across the semester and many were able to adapt to unplanned interactions during lesson delivery. Boody [19] referred to these in-the-moment adaptations as reflection-in-action. Coaching also encouraged the treatment participants to reflect critically about how they performed a lesson [19]. In contrast, the comparison group participants did not report that they engaged in critical reflection.

9. Conclusions, Limitations, and Future Research

We explored the relationship between Bandura's [14,15] theory of self-efficacy, data-driven feedback and coaching strategies [9–11], and Boody's [19] reflective practices in our mixed-methods design.

Based on the self-efficacy theory, our theoretical framework did not lead to a statistically significant difference between groups on the mean TSES scores but did have significance for one target item about question development. A single semester, three-fourths of the way through a teacher education program, might not have been the best time to measure change in general perceptions of self-efficacy. A pretest and posttest analysis extending over the entire program would provide a better perspective for change. In addition, the high overall scores reported for both groups may be due to their participation in the entirety of the education program rather than the effects of the treatment in one course. When candidates are exposed to best practices in teaching over an extended period, such as two to three years, a clear measure of improvement in self-efficacy can be traced, as reported by Gundel et al. [27].

What did we learn about feedback and coaching? The data-driven feedback used in this study was effective because it could be defined and operationalized. It directly related to the task of improving HOT questioning skills and the data were easily derived from observation. Use of specific feedback made the process clear for the researcher and the candidate. The feedback had immediate meaning to the learners who could use it right away within the coaching session to initiate a dialogue about how to improve [59]. Unfortunately, the study could not be extended to ascertain if coached individuals improved their use of HOT skills beyond the semester.

The small sample size for this study made it less generalizable, yet more manageable to review each video, obtain feedback data, contact participants, and provide effective coaching. Future researchers may consider a sample size of 15 as a target caseload for a single researcher on a team. Although lessons were limited to 10 min, the notion of

an optimal ratio of two K/C to one HOTS question in this timeframe is a concept that could be tested by other researchers to develop expectations for the demonstration of questioning skills.

Discriminating between the three levels of HOTS questions evolved after all sessions were completed. As we listened to the video recordings, we realized that some HOT questions engaged the students much more than others. We soon asked ourselves about the quality of the HOT questions and distinguished three features: higher-order thinking, wait-time, and student engagement. HOT questions combining all three criteria were designated as having the highest quality, however, they also took longer to initiate, limiting the number of HOT questions that could be posed in our timeframe. We also analyzed the progression of HOT questions across the three sessions, as seen in Table 4, noting that there were no level 3 questions in session 1. Once candidates became confident in asking level 1 HOT questions, they added more wait-time to get a response. When students began to interact with each other, the preservice teachers were surprised that the HOT questions encouraged critical thinking and reduced classroom management issues. Directly teaching students about the different levels of HOT questions and coaching them to reach a level 3 could be a future research activity. Providing feedback about the exact level of the HOT question with respect to the cognitive domain could also be used to refine questioning techniques. Varying the amount of wait-time after a question is posed could also be examined. Additionally, student responses to questions could be recorded and coded to investigate the dialogue and synergy that occurs when students are engaged.

These teacher candidates responded that they felt supported by the feedback and coaching model we designed within the MRS environment. Not only did the treatment participants improve their use of HOT questions, but they also applied multiple types of self-reflection [19] and positively impacted their perceptions of self-efficacy [33] as indicated in their personal accounts. The MRS environment facilitated a consistent classroom setting for lesson delivery and data collection. Candidates easily referred to the avatars as “my students” or “my class”, reinforcing the value of MRS as a viable replacement for in-person teaching to gain initial skills as a classroom teacher.

Author Contributions: Conceptualization, W.J.D., M.A.B.D., B.M.S. and J.C.G.; methodology, W.J.D., M.A.B.D., B.M.S. and J.C.G.; validation, W.J.D. and M.A.B.D.; formal analysis, W.J.D. and M.A.B.D.; writing—original draft preparation, W.J.D. and M.A.B.D.; writing—review and editing, W.J.D., M.A.B.D., B.M.S. and J.C.G.; project administration, W.J.D., M.A.B.D., B.M.S. and J.C.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Western Connecticut State University (protocol code: 1617-95, date of approval: 21 December 2016).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are not available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Forzani, F.M. Understanding “core practices” and “practice-based” teacher education: Learning from the past. *J. Teach. Educ.* **2014**, *65*, 357–368. [\[CrossRef\]](#)
2. Aulls, M.W.; Shore, B.M. *Inquiry in Education: The Conceptual Foundations for Research as a Curricular Imperative*; Taylor & Francis Group: New York, NY, USA, 2008.
3. Shore, B.M.; Aulls, M.W.; Delcourt, M.A.B. (Eds.) *Inquiry in Education: Overcoming Barriers to Successful Implementation*; Taylor & Francis Group: New York, NY, USA, 2008.
4. Salinas, C.; Blevins, B. Critical historical inquiry: How might preservice teachers confront master historical narratives? *Soc. Stud. Res. Pract.* **2014**, *9*, 35–50. [\[CrossRef\]](#)
5. Bloom, B.S. *Taxonomy of Educational Objectives, Book 1: Cognitive Domain*, 2nd ed.; Longman: Ann Arbor, MI, USA, 1956.

6. Anderson, L.W.; Krathwohl, D.R. (Eds.) *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*; Longman: New York, NY, USA, 2001.
7. Goodwin, A.L.; Smith, L.; Souto-Manning, M.; Cheruvu, R.; Tan, M.Y.; Reed, R.; Taveras, L. What should teacher educators know and be able to do? Perspectives from practicing teacher educators. *J. Teach. Educ.* **2014**, *65*, 284–302. [CrossRef]
8. McDonald, M.; Kazemi, E.; Kavanagh, S.S. Core practices and pedagogies of teacher education: A call for a common language and collective activity. *J. Teach. Educ.* **2013**, *64*, 378–386. [CrossRef]
9. Hattie, J.; Biggs, J.; Purdie, N. Effects of learning skills interventions on student learning: A meta-analysis. *Rev. Educ. Res.* **1996**, *66*, 99–136. [CrossRef]
10. Hattie, J.; Timperley, H. The power of feedback. *Rev. Educ. Res.* **2007**, *77*, 81–112. [CrossRef]
11. Wisniewski, B.; Zierer, K.; Hattie, J. The power of feedback revisited: A meta-analysis of educational feedback research. *Front. Psychol.* **2020**, *10*, 1–14. [CrossRef]
12. Blackley, S.; Sheffield, R.; Maynard, N.; Koul, R.; Walker, R. Makerspace and reflective practice: Advancing preservice teachers in STEM education. *Aust. J. Teach. Educ.* **2017**, *42*, 22–37. [CrossRef]
13. Kraft, M.A.; Blazar, D.L. Individualized coaching to improve teacher practice across grades and subjects: New experimental evidence. *Educ. Policy* **2017**, *31*, 1033–1068. [CrossRef]
14. Bandura, A. *Social Foundations of Thought and Action*; Prentice Hall: Englewood Cliffs, NJ, USA, 1986.
15. Bandura, A. Self-efficacy. In *Encyclopedia of Human Behavior*; Ramachaudran, V.S., Ed.; Academic Press: Cambridge, MA, USA, 1994; Volume 4, pp. 71–81.
16. Bandura, A. *Self-Efficacy: The Exercise of Control*; Freeman: New York, NY, USA, 1997.
17. Averill, R.; Drake, M.; Harvey, R. Coaching pre-service teachers for teaching mathematics: The views of students. In *Mathematics Education: Yesterday, Today, and Tomorrow*; Steinle, V., Ball, L., Bordini, C., Eds.; Mathematics Education Research Group of Australasia (MERGA): Payneham, Australia; pp. 707–710.
18. Bandura, A. Human agency in social cognitive theory. *Am. Psychol.* **1989**, *44*, 1175–1184. [CrossRef]
19. Boody, R.M. Teacher reflection as teacher change, and teacher change as moral response. *Education* **2008**, *128*, 498–506.
20. Killion, J.P.; Todnem, G.R. A process for personal theory building. *Educ. Leadersh.* **1991**, *48*, 14–16.
21. Liu, K. Critical reflection as a framework for transformative learning in teacher education. *Educ. Rev.* **2015**, *67*, 135–157. [CrossRef]
22. Reagan, T.G.; Case, C.W.; Brubacher, W. *Becoming a Reflective Educator: How to Build a Culture of Inquiry in the Schools*; Corwin Press: Thousand Oaks, CA, USA, 2000.
23. Dieker, L.A.; Rodriguez, J.A.; Lignugaris/Kraft, B.; Hynes, M.C.; Hughes, C.E. The potential of simulated environments in teacher education. *Teach. Educ. Spec. Educ. J. Teach. Educ. Div. Counc. Except. Child.* **2013**, *37*, 21–33. [CrossRef]
24. Dieker, L.A.; Straub, C.L.; Hughes, C.E.; Hynes, M.C.; Hardin, S. Learning from virtual students. *Educ. Leadersh.* **2014**, *71*, 54–58. Available online: <https://www.learntechlib.org/p/153594/> (accessed on 11 April 2023).
25. Anton, S.; Piro, J.; Delcourt, M.A.B.; Gundel, E. Pre-service teacher's coping and anxiety within mixed reality simulations. *Soc. Sci.* **2023**, *12*, 146. [CrossRef]
26. Cohen, J.; Wong, V.; Krishnamachari, A.; Berlin, R. Teacher coaching in a simulated environment. *Educational Evaluation and Policy Analysis* **2020**, *42*, 208–231. [CrossRef]
27. Gundel, E.; Piro, J.; Straub, C.; Smith, K. Self-efficacy in mixed reality simulations: Implications for pre-service teacher education. *Teach. Educ.* **2019**, *54*, 244–269. [CrossRef]
28. Howell, H.; Mikeska, J.N.; Croft, A.J. *Simulations Support Teachers in Learning How to Facilitate Discussions*; American Educational Research Association: Toronto, ON, Canada, 2019. Available online: https://convention2.allacademic.com/one/aera/aera19/index.php?cmd=Online+Program+View+Paper&selected_paper_id=1432554&PHPSESSID=817shaqo7a8apbg15pa1v4ghel (accessed on 11 April 2023).
29. Mendelson, E.; Piro, J. An affective, formative, and data-driven feedback intervention in teacher education. *Curric. Teach.* **2022**, *8*, 13–29. [CrossRef]
30. Piro, J.; O'Callaghan, C. Journeying towards the profession: Exploring liminal learning within mixed reality simulations. *Action Teach. Educ.* **2018**, *41*, 79–95. [CrossRef]
31. Rosati-Peterson, G.; Piro, J.; Straub, C.; O'Callaghan, C. A nonverbal immediacy treatment with pre-service teachers using mixed reality simulations. *Cogent Educ.* **2021**, *8*, 2–39. [CrossRef]
32. Mursion, V.R. About Us. Available online: <https://www.mursion.com/services/education/> (accessed on 11 April 2023).
33. Bandura, A. Social cognitive theory: An agentic perspective. *Annu. Rev. Psychol.* **2001**, *52*, 1–26. [CrossRef] [PubMed]
34. Bandura, A. Perceived self-efficacy in cognitive development and functioning. *Educ. Psychol.* **1993**, *28*, 117–148. [CrossRef]
35. Üstünbaş, Ü.; Alagöz, N. Efficacy beliefs and metacognitive awareness in English language teaching and teacher education. *Bartın Univ. J. Fac. Educ.* **2021**, *10*, 267–280. [CrossRef]
36. Hargrove, T.; Fox, K.R.; Walker, B. Making a difference for pre-service teachers through authentic experiences and reflection. *Southeast. Teach. Educ. J.* **2010**, *3*, 45–54.
37. Dewey, J. *How We Think*; Heath: New York, NY, USA, 1933.
38. Schön, D. *The Reflective Practitioner*; Basic Books: New York, NY, USA, 1983.

39. Aulls, M.W. Developing students' inquiry strategies: A case study of teaching history in the middle grades. In *Inquiry in Education: Overcoming Barriers to Successful Implementation*; Shore, B.M., Aulls, M.W., Delcourt, M.A.B., Eds.; Taylor & Francis Group: New York, NY, USA, 2008; pp. 1–46.
40. Bramwell-Rejskind, F.G.; Halliday, F.; McBride, J.B. Creating change: Teachers' reflections on introducing inquiry teaching strategies. In *Inquiry in Education: Overcoming Barriers to Successful Implementation*; Shore, B.M., Aulls, M.W., Delcourt, M.A.B., Eds.; Taylor & Francis Group: New York, NY, USA, 2008; pp. 207–234.
41. Robinson, A.; Hall, J. Teacher Models of Teaching Inquiry. In *Inquiry in Education: Overcoming Barriers to Successful Implementation*; Shore, B.M., Aulls, M.W., Delcourt, M.A.B., Eds.; Taylor & Francis Group: New York, NY, USA, 2008; pp. 235–246.
42. Danielson, C. *The Framework for Teaching: Evaluation Instrument*; The Danielson Group: Chicago, IL, USA, 2013.
43. Delcourt, M.A.B.; McKinnon, J. Tools for inquiry: Improving questioning in the classroom. *Learn. Landsc.* **2011**, *4*, 145–159. [\[CrossRef\]](#)
44. Dieker, L.A.; Hughes, C.E.; Hynes, M.C.; Straub, C. Using simulated virtual environments to improve teacher performance. *Sch.-Univ. Partnersh* **2017**, *10*, 62–81.
45. Kraft, M.A.; Blazar, D.; Hogan, D. The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Rev. Educ. Res.* **2018**, *88*, 547–588. [\[CrossRef\]](#)
46. Joyce, B.R.; Showers, B. Transfer of training: The contribution of “coaching”. *J. Educ.* **1981**, *163*, 163–172. [\[CrossRef\]](#)
47. Khalil, D.; Gosselin, C.; Hughes, G.; Edwards, L. Teachlive™ Rehearsals: One HBCU's study on prospective teachers' reformed instructional practices and their mathematical affect. In Proceedings of the 38th Annual Meeting of the North American Chapter of the International Group of the Psychology of Mathematics Education, Tucson, AZ, USA, 3–6 November 2016; pp. 767–774.
48. Garland, K.V.; Vasquez, E., III; Pearl, C. Efficacy of individualized clinical coaching in a virtual reality classroom for increasing teachers' fidelity of implementation of discrete trial teaching. *Educ. Train. Autism Dev. Disabil.* **2012**, *47*, 502–515. Available online: <https://www.jstor.org/stable/23879642> (accessed on 11 April 2023).
49. Pas, E.T.; Johnson, S.R.; Larson, K.E.; Brandenburg, L.; Church, R.; Bradshaw, C.P. Reducing behavior problems among students with autism spectrum disorder: Coaching teachers in a mixed-reality setting. *J. Autism Dev. Disord.* **2016**, *46*, 3640–3652. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Gall, M.; Gall, J.; Borg, W. *Educational Research: An Introduction*, 8th ed.; Pearson: New York, NY, USA, 2006.
51. Creswell, J.W.; Plano Clark, V.L. *Designing and Conducting Mixed Methods Research*, 2nd ed.; SAGE: Thousand Oaks, CA, USA, 2011.
52. Tschannen-Moran, M.; Hoy, A.W. Teacher efficacy: Capturing an elusive construct. *Teach. Teach. Educ.* **2001**, *17*, 783–805. [\[CrossRef\]](#)
53. CAEP. Council for Accreditation of Educator Preparation. 2023. Available online: <https://caepnet.org> (accessed on 11 April 2023).
54. Nie, Y.; Lau, S.; Liau, A. The Teacher Efficacy Scale: A reliability and validity study. *Asia Pac. J. Educ.* **2012**, *21*, 414–421. Available online: <http://hdl.handle.net/10497/14287> (accessed on 11 April 2023).
55. Westberg, K.; Archambault, F.; Dobyns, S.; Salvin, T. *An Observational Study of Instructional and Curricular Practices Used with Gifted and Talented Students in Regular Classrooms*; Research Report No. 93104; The National Research Center on the Gifted and Talented: Storrs, CT, USA, 1993.
56. Hinkle, D.E.; Wiersma, W.; Jurs, S.G. *Applied Statistics for the Behavioral Sciences*; Houghton Mifflin: Boston, MA, USA, 2003.
57. Saldaña, J. *The Coding Manual for Qualitative Researchers*; SAGE: Thousand Oaks, CA, USA, 2016.
58. Laverick, V.T. Secondary teachers' understanding and use of reflection: An exploratory study. *Am. Second. Educ.* **2017**, *45*, 56–68. Available online: <https://www.jstor.org/stable/45147895> (accessed on 11 April 2023).
59. Adie, L.; van der Kleij, F.; Cumming, J. The development and application of coding frameworks to explore dialogic feedback interactions and self-regulated learning. *Br. Educ. Res. J.* **2018**, *44*, 704–723. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.