Year 1 Findings
University of Central Florida
Carrie Straub, Ph.D.
Lisa Dieker, Ph.D.
Michael Hynes, Ph.D.
Charles Hughes, Ph.D.
Using Virtual Rehearsal in TLE TeachLivE™ Mixed Reality Classroom Simulator to Determine the Effects on the Performance of Mathematics Teachers

This research project was supported by funding from the Bill & Melinda Gates Foundation.

Abstract

A quasi-experimental, pre-post group design was used to examine the effects of repeated virtual rehearsal sessions in a mixed-reality computer simulated environment, the TLE TeachLivE™ classroom simulator. At 10 sites across the nation, 157 middle school mathematics teachers received four levels of innovative professional development, including computer simulation, synchronous online instruction, and lesson resources based on the Common Core State Standards (CCSS). Teachers were observed in the TLE TeachLivE™ (TeachLivE) classroom simulator and in their regular classrooms to determine the effects of treatment. Trained observers (a) collected pre-post frequency counts of teacher behavior on questioning, wait time, and feedback; (b) scored teacher classroom practice on modified sub-constructs of the Danielson Framework for Teaching; and (c) took qualitative field notes. Teachers responded to questionnaires covering demographic information and experiences in the classroom simulator. Student level data were collected pre-post on a 10-item academic assessment using items from the eighth grade 2011 National Assessment of Educational Progress. From an analysis of the data, the researchers found that four 10-minute professional learning sessions in the TeachLivE classroom simulator improved targeted teaching behaviors in the simulator scenarios, and those improvements transferred into the teachers’ original classroom settings. Results from this study validate emerging research in the field of teacher education and simulation that suggests that professional learning in mixed-reality simulated classrooms can be effective in impacting teacher practice.
Using Virtual Rehearsal in TLE TeachLivETM Mixed Reality Classroom Simulator to Determine the Effects on the Performance of Mathematics Teachers

Teachers are the single most important factor to influence student learning and academic outcomes, aside from the students themselves (Darling-Hammond, 2003; Hattie, 2003; Kane & Staiger, 2008). Since 2009, President Barack Obama’s administration has invested over $4 billion in four key reform areas, one of which includes a focus on helping teachers become more effective. In the White House’s 2011 Strategy for American Innovation, citizens of the United States were called on to “out-innovate, out-educate, and out-build the rest of the world” (National Economic Council, 2011, p.1) to maintain America’s economic and competitive growth. The administration’s strategy emphasizes improving teacher preparation for science, technology, engineering, and mathematics (STEM) subjects.

U.S. students continue to struggle in STEM areas, with scores not commensurate of a world power (U.S. Department of Education, 2008). New standards are in place to support teachers (Common Core Standards Initiative [CCSI], 2011), and educational stakeholders have called for an overhaul of the Elementary and Secondary Education Act, the largest piece of legislation relating to education in the U.S., to address these issues. Though student scores on the National Assessment of Educational Progress (NAEP) are improving in fourth and eighth grades, only 35% of eighth grade students reached or exceeded proficiency in 2013 (National Center for Education Statistics, 2013). National supporters of education reform call for improvements that include reshaping of teachers’ pedagogical and content knowledge in targeted areas. Skilled pedagogy and delivery of content knowledge are required in order to ensure maximized teacher influence on learning in all content areas within these new standards. As more states adopt standards aligned to the CCSI, coupled with increased rigor in mathematics, students will experience academic challenges prior to graduation. Likewise, the challenge is passed on to teachers, who are expected to prepare these students for college and careers in a more advanced STEM society.

If teachers are one of the key factors to students’ academic success, what characterizes typical teacher professional development (PD) in STEM areas? In a national sample of over 1,000 mathematics and science teachers, Garet and colleagues (2001) found three key features of PD related to changes in classroom practices and self-reported increases in knowledge and skills: (a) focus on content knowledge; (b) opportunities for active learning; and (c) coherence with other learning activities. In findings from two longitudinal studies, the National Longitudinal Study of No Child Left Behind and the Study of State Implementation of Accountability and Teacher Quality under No Child Left Behind, Birman and colleagues (2007) reported that only 8% of elementary, middle, and high school teachers received PD on instructional strategies in mathematics. Further, only 16% of secondary mathematics teachers reported participation in extended PD (over 24 hours in one year) on instructional strategies specific to mathematics. Cost and time are two challenges to providing effective PD (Garet et al., 2001; Guskey & Yoon, 2009). Birman et al. (2007) reported that teachers participated in an average of 66 hours of PD during one calendar year, yet “less than one-quarter of teachers reported that they participated in professional development that often provided opportunities to practice what they had learned, lead discussions, or conduct demonstrations” (p. 76).
High quality PD is crucial for teachers to meet the new levels of learning standards in place today, whether a state has or has not adopted the CCSS. The ultimate outcome of any PD is determining the impact on student academic outcomes (Loucks-Horsley & Matsumoto, 1999). Due to the complex nature of collecting student data in schools (Guskey & Sparks, 2002; Loucks-Horsley & Matsumoto, 1999), there is limited research meeting the What Works Clearinghouse (WWC) standards for evaluating the impact of PD on student achievement (Guskey & Yoon, 2009; U.S. Department of Education, 2008; Yoon, Duncan, Wen- Yu Lee, Scarloss, & Shapley, 2007). Yoon and colleagues (2007) reported a lack of rigorous research regarding the effects of teacher PD on student achievement, identifying over 1,300 studies between 1986 and 2003 of which only 9 met the WWC evidence standards and all were at the elementary school level. In a follow-up analysis conducted by Guskey and Yoon (2009), each of the nine studies cited active learning and opportunities for teachers to adapt practices to their individual classrooms as having the greatest impact.

Desimone, Porter, Garet, Yoon, and Birman (2002) conducted a longitudinal study of 207 math and science teachers in 30 schools in five states and found that PD that included active learning opportunities increased the effect on teachers’ instruction. Active learning is defined by Desimone and colleagues as “opportunities for teachers to become actively engaged in the meaningful analysis of teaching and learning, for example, by reviewing student work or obtaining feedback on their teaching” (p. 83). They also found that PD with a focus on specific teaching practices predicted increased use of these practices in the classroom.

Teachers engage in an array of specific teaching practices in their classrooms, and PD should target the practices teachers find the most challenging. In the Measures of Effective Teaching study, Kane and Staiger (2012) reported that teachers scored the highest for competencies related to creating an orderly environment, and lowest for complex teaching skills such as questioning, discussion techniques, and communicating with students about content. As Shulman (1986) posited, teaching requires strong pedagogical content knowledge, an understanding of not only subject matter but also the in-depth approaches for teaching the subject matter. Some practices span the content areas and focus on aspects of teaching that teachers may find challenging, but have the greatest rewards for students. Teaching Works (2014) analyzed core capabilities for teachers and developed a set of nineteen high-leverage practices (HLPs) for teaching across content areas, including STEM subjects. Mastering these practices will likely lead to increased advances in student learning. The practices are based on research linking particular practices to student achievement and are generated from published descriptions of teaching, videos of teachers at work, and expert experience (Loewenberg Ball & Forzani, 2010). The Teaching Works HLPs span across content, teacher style, and setting, and include practices such as eliciting and interpreting student thinking, and providing oral feedback on students’ work (Loewenberg Ball, Sleep, Boerst, & Bass, 2009).

Similar teaching capabilities are described in other published descriptions of teacher practice. Danielson (2011) provided indicators for eliciting student thinking, such as higher-level questioning. Higher-level questions are defined as open-ended questions that allow students to use past experiences, prior knowledge, and previously learned content and relate it to newly learned content in order to create a well thought-out answer (i.e. question statements that begin with “How”, “What”, or “Why”). Danielson advocates that after teachers are asked
higher-level questions, they should provide students with sufficient time to think about their responses, reflect on the comments of their classmates, and deepen their understanding.

If research is converging on a core set of high-quality teaching practices that positively impact student outcomes, and researchers have identified characteristics of high-quality PD for teachers, what are the best environments for delivering PD to teachers? Next generation PD environments for teachers to learn both pedagogical and content skills are emerging, and computer simulation is at the forefront. Dieker, Straub, Hughes, Hynes, and Hardin (2014) described simulated environments divided into four levels based on complexity and evolution of technology. Levels one and two of simulated environments can be found in the current education landscape, and levels three and four are in various stages of development: (1) virtual reality desktop: users interact with avatars (virtual characters) using a typical office computer and display monitor with a mouse and keyboard for operation; (2) mixed-reality: real and virtual worlds are combined, giving users a sense of presence and immersions, with varying types of displays such as a large monitor, rear-projection screen, or head-mounted display combined with a tracking device for user movement; (3) immersive 3-D environments: emerging technologies in which users interact with avatars that have moved from the virtual to the real space and are able to physically interact with users; and (4) brain-computer interfaces: in the future, technology will allow users to remotely interact with their environment through their senses. At all levels, users interact with computer simulations that blend synthetic content with reality but continue to work on targeted skills. These simulated environments can provide many educational experiences and opportunities that may not be available in real-world settings (Dieker, Rodriguez, Lignugaris/Kraft, Hynes, & Hughes, 2014; Nagendran, Pillat, Kavanaugh, Welch, & Hughes, 2014) and allow for safe rehearsal of skills until mastery.

**Virtual Reality Desktop Simulations**

Virtual reality desktop classroom simulations have been evaluated with pre-service and in-service teachers on a variety of dimensions, using a variety of software platforms (e.g., PLATO, simSchool, and Cook County School District). Boyce (1988) described one of the first desktop simulated classrooms (PLATO), “A Simulation of the First Year Teaching” (Gaedes, 1975). Teacher candidates were given the objective of earning tenure from their principals and made decisions related to student seating, class rules, grading policies, discipline problems, club sponsorship, teachers’ unions, and team teaching. Users reported new awareness of the complexities of teaching, and many users repeated the simulation more than once, which led the researchers to conclude that users found it meaningful and fun.

Foley and McAllister (2005) evaluated desktop classroom simulation (i.e., simSchool) designed to create a reality-based context for teacher candidates to practice teaching skills, reflect on decision-making, and draw connections from theory to practice. A teacher efficacy scale was used to quantitatively compare teacher candidates’ pretest to posttest scores with no comparison group using paired sample t-tests for each of the four constructs: (a) instructional efficacy, (b) disciplinary efficacy, (c) ability to create a positive school climate, and (d) ability to enlist community. Researchers reported small significant increases in perceptions of disciplinary efficacy and ability to create a positive school climate, although descriptive and analytic statistical results were not provided, and low response rates were reported due to coding
problems between pretests and posttests. Lack of increased efficacy for instruction was attributed to candidates’ increased awareness of complexity of teaching as a result of the simulation, the implication being that while the simulation made candidates aware of the needs of diverse learners, it had not resulted in increased self-efficacy to address those needs.

McPherson, Tyler-Wood, McEnturff, and Peak (2011) investigated the effects of desktop classroom simulation software (i.e., simSchool) in a pre-post quasi-experimental, non-equivalent comparison group design to explore perceptions of 151 teachers and teacher candidates in terms of preparedness for teaching and inclusion of students with disabilities. The findings indicated significantly higher effects for teaching skills for both graduate ($t(42) = 2.72$, $p = .01$) and undergraduate ($t(23) = 2.26$, $p = .03$) level students, and for instructional self-efficacy for undergraduate students ($t(23) = -3.90$, $p < .001$). However, no significant effects were found for attitudes toward inclusion. Researchers indicated that graduate students may have had teaching experience, which would give them more realistic perceptions of their instructional efficacy; therefore their test scores would not change significantly from pretest to posttest, while undergraduates with no teaching experience may tend to overestimate their efficacy at pretest.

Girod and Girod (2006) evaluated six hours of desktop classroom simulation software (i.e., Cook County School District) with 71 teacher candidates in a quasi-experimental comparison group study. An additive effect was found for the treatment on work sample scores of teacher candidates ($F(1,5) = 4.56$, $p < .05$, $\eta^2_p = .48$). Work samples are a proxy for performance in a real classroom, so observers also conducted lesson plan evaluations in teacher candidates’ field placement classrooms on five dimensions: (a) provided evidence of planning for instruction, (b) established a classroom climate conducive to learning, (c) implemented plans for instruction, (d) evaluated pupil achievement, and (e) demonstrated an impact on student learning. Additive effects were found for planning for instruction ($F(1,67) = 4.36$, $p < .05$, $\eta^2_p = .06$) and establishing a classroom conducive to learning ($F(1,37) = 8.15$, $p < .05$, $\eta^2_p = .18$), but significant differences were not found for the other dimensions. While ability to evaluate pupil achievement approached significance, plans for instruction and impact on student learning were not found. Researchers concluded that study results aligned with the focus of the simulation, which placed “great emphasis on evaluating pupil achievement and none on actual, real-world teaching or implementing plans for instruction” (p. 492). While Girod and Girod included practicing teachers in their participant pool, results were not aggregated by status of teacher or teacher candidate.

Simulated environments provide a safe place to practice teaching behaviors at an accelerated pace and receive rapid corrective feedback (Dieker et al., 2014a; McPherson et al., 2011). However, in quasi-experimental simulation studies noted above, researchers found effects for efficacy and work products for planning instruction, but found less evidence to suggest change in instructional practice in the classroom as a result of simulation. Because instruction is at the heart of teaching, and mastering skills needed for pedagogical content knowledge is challenging for teachers (Kane & Staiger, 2012; Shulman, 1986), the use of simulation should be explored to recreate scenarios that give teachers and teacher candidates effective practice in instruction. Perhaps simulations with higher fidelity that resemble the act of teaching are needed to create significant effects for users, much like in an immersive flight simulator, when pilots are closer to the real act of flying to master targeted skills. In a similar fashion, teachers could practice in a classroom simulator, which might have higher fidelity aligned to the true act of teaching. While
desktop simulations provide users with opportunities to make teaching-related decisions, a more authentic simulation of teaching might result from a mixed-reality environment.

**Mixed-reality Simulations**

TLE TeachLivE™ (TeachLivE) is an immersive, mixed-reality classroom simulator that includes the features of real classroom with desks, teaching materials, whiteboards, and students. Real and virtual worlds are combined to give users a sense of immersion and presence, and the teachers interact with student-avatars in real time, holding authentic discussions on varied content areas. Student-avatars have personalities typical of real-life students, and teachers are faced with instructional decisions based on varying levels of behavioral compliance and content knowledge, much like in a real classroom. Over 40 university or school district partners currently have TeachLivE classroom simulators for teacher professional learning, and TeachLivE is currently the only mixed-reality classroom simulator of its kind. A research base, focusing on the use of TeachLivE with teachers and teacher candidates, is emerging, and TeachLivE is currently in its fourth generation of student-avatars.

Early research conducted by Andreasen and Haciomeroglu (2009), using a mixed methods study with 15 teacher candidates, investigated TeachLivE as part of a mathematics methods course. Candidates were divided into five groups, and each group rotated through a three-stage cycle of teaching in which one member taught the lesson while the other two observed. All 15- to 20-minute sessions were videotaped, featuring middle school student-avatars with designated work samples for the simulation. After completing the simulation, candidates wrote reflections on their performance. During the next class, they watched videos of their interactions and revised their lessons for the next TeachLivE simulation. Qualitative observational data were reported. Teachers experienced student-avatar misbehavior and the need to balance content delivery with behavior management, including challenges to teaching authority. “The focus of the lesson became managing behavior and the content was left to the wayside” (p. 1322). Researchers concluded that the simulated classroom allowed for a focus on managing student behavior in order for delivery of content to occur—a unique capability over other teacher training environments, such as micro-teach or role-play, in which students practice delivering content to a classroom of their peers.

Four teachers took part in six 15-minute TeachLivE sessions to learn discrete trial teaching, an evidence-based practice for students with autism (Vince Garland, Vasquez, & Pearl, 2012). Researchers used a multiple baseline design to evaluate teachers’ fidelity of implementation. Overall mean accuracy of implementing discrete trial teaching increased from 37% in baseline to 87% in treatment after six 15-minute sessions in the TeachLivE classroom simulator. All teachers reported that they would use knowledge gained during the simulation in their classrooms. Two of the four teachers reported they were more comfortable making mistakes and learning in front of student-avatars than real students. One teacher reported she felt learning in TeachLivE was more efficient, because the opportunities to respond were guaranteed, as student-avatars’ behaviors could be controlled.

Eleven teachers and teacher candidates used the TeachLivE classroom simulator to practice a read-aloud activity during one 5-minute session followed by after-action-review
(Elford, James, & Haynes-Smith, 2013). Qualitative data were collected from semi-structured interviews, and three themes emerged: (a) classroom management skills were needed to successfully complete the simulation, as teachers reported the student-avatars had realistic personalities when compared to their own students; (b) although teachers were initially distracted by student-avatars’ jerky movements (Generation 1 TeachLivE avatars, used in this study, were an early prototype with skeletal movement issues), they were able to look past this and interact with the student-avatars as if they would real students; and (c) TeachLivE provided a valuable opportunity for teachers to reflect on their instruction and determine how to improve.

Elford, Carter, and Aronin (2013) used TeachLivE as a practice environment to give four secondary teachers feedback on their performance addressing student behaviors. All four teachers wore a Bluetooth device that allowed them to hear real-time prompts from an expert coach during their simulation sessions. Participants took part in four 5-minute sessions, half of which were randomly selected for teachers to receive prompts via Bluetooth, with no prompts in the remaining sessions, such that all teachers received prompts half of the time. When participants received remote coaching via Bluetooth, the percentage of addressed behaviors increased, with positive feedback increasing from 20% to 30% across all participants and all sessions. Teachers reported increased comfort interacting with the student-avatars as if they were real students. One teacher reported, “…getting this kind of practice is so much more meaningful than just listening to someone talk about how to do a certain strategy” (p. 43).

Whitten, Enicks, Wallace, and Morgan (2013) conducted a two-group randomized design experiment with teacher candidates over four 10-minute TeachLivE sessions throughout one academic year. TeachLivE was compared to online modules, and treatments were counterbalanced so that all participants had access to both conditions. A classroom observation tool was used to evaluate candidates and provide opportunity for specific feedback. Results indicated that candidates’ mean scores increased over time, yet results could not be attributed to TeachLivE interventions, and comparison group data were not provided. Researchers recommended integrating simulation sessions across the program, so that candidates who were not meeting minimum program requirements would receive targeted experiences in TeachLivE at increasing levels of intensity until skill mastery was demonstrated.

Dawson and Lignugaris/Kraft (2013) compared practice sessions in role-play (a traditional approach to classroom simulation) to practice sessions in TeachLivE (a technology-enhanced approach to classroom simulation) on teachers’ utilization of four evidence-based strategies for teaching: (a) opportunities to respond, (b) praise, (c) error correction, and (d) praise around. These researchers used an innovative alternating treatments design for two consecutive studies with seven teachers split into two groups, using a counterbalanced design across treatments and groups. For both studies, in order to evaluate generalization of skills, intervention was delivered in Generation 1 TeachLivE, and teachers were assessed in Generation 3 TeachLivE. Teachers in both studies had a higher response rate for the skill practiced in TeachLivE than the skill practiced in role-play; however, performance levels were similar for both groups at the close of study two, leading researchers to question the extent to which these differences would maintain across time.
A mixed-reality classroom simulation has the potential to deliver targeted learning activities as PD for teachers. One of the purposes of this research study is to evaluate the use of a classroom simulator with high fidelity, TeachLivE, to affect actual classroom instruction. In this study, we proposed to provide teachers an opportunity to practice their use of HLPs such as higher-level questioning, specific praise, and wait time in TeachLivE and to evaluate the generalization of those practices to the traditional classroom setting. High fidelity simulators have unique capabilities that make them advantageous over observing and providing feedback in a typical classroom.

One such advantage is the capability for use of integrated video tagging software to record, play back, and export data collected during a simulation session, without the need of prior approval from students’ parents. Observers and the software collect information about teacher practice that can be used for data analysis at a later date, increasing reliability of observations. Teachers can review their performance immediately after teaching, by pausing the simulation while the experience is fresh and either reviewing their practice by video or discussing their performance with an expert coach. In this scenario, teachers receive just-in-time PD, and real students are not made to wait while their teacher receives corrective feedback. The simulated activity allows time for an integrated after-action-review process to take place, in which teachers take part in structured reflection (Baird, Holland, & Deacon, 1999). Most importantly, unlike in real classrooms, teachers can re-enter the environment to fix instructional errors with student-avatars without affecting real students. Potentially, immersive virtual environments can change the face of teacher PD with innovative applications of the technology, but research is needed to establish the efficacy of the use of simulation for teacher education.

**Theoretical Framework and Overarching Hypotheses**

Based on results from earlier studies related to using virtual environments for teacher preparation, our overarching hypothesis is that teachers who engage in virtual environment simulations will improve their pedagogical knowledge as well as student content knowledge. We developed a theory of action based on relevant simulation research related to professional learning, after examining the features of professional learning and their relationship to teacher practice and student outcomes (e.g., active learning opportunities based on specific teaching practices, such as HLPs).

We hypothesized that teacher learning is most effective in contextually meaningful settings (Dieker et al., 2014a), and created a contextually meaningful simulation activity that provided learners with the opportunity to practice HLPs with student-avatars. Our work was grounded in Brown, Collins, and Duguid’s (1989) theory of situated cognition asserting that “what is learned cannot be separated from how it is learned and used” (p. 88). Further, we hypothesized that learning that occurred in a virtual classroom would transfer to a real classroom. Specifically, we hypothesized that four 10-minute sessions of virtual rehearsal (i.e., practicing the same lesson and HLPs in TeachLivE) would increase teachers’ frequency of higher order questions and specific feedback to students. Furthermore, we hypothesized that this increase of two effective teaching practices would be observed during both simulated and real classroom instruction. Also, we hypothesized that online PD by itself would increase teachers’
frequency of practices and, when combined with TeachLivE, a differential increase would be evident.

**Research Questions for Teacher Performance**

As outlined above, the focus of this research study was on changing teacher practice. We set about finding evidence of change in teacher practice in two environments: (a) the classroom simulator and (b) the teachers’ classrooms. In both settings we attempted to change teacher practice using TeachLivE or TeachLivE combined with other forms of PD. Our first set of research questions and hypotheses focused on the classroom simulator environment, and the second set focused on the classroom. In the TeachLivE classroom simulator, we had the following research questions:

*Research Questions 1 and 2:* Are there differences in performance over four 10-minute sessions of TeachLivE in a classroom simulator based on whether or not teachers received 40 minutes of online PD?

*Question 1:* on frequency of describe/explain questions asked during a 10-minute simulation?

*Question 2:* on frequency of specific feedback given during a 10-minute simulation?

Ultimately, teacher practice takes place in a classroom setting, so it was important to identify whether or not there was evidence of an effect in a classroom with students present. Simulation generally incorporates an after-action-review process (Smith & Allen, 1994) that gives the user feedback on performance. We hypothesized withholding feedback on a specific teacher practice (i.e., wait time) would result in no differences in wait time across groups, whether teachers received four 10-minute sessions in TeachLivE or not.

*Research Question 3:* What are the effects of simulation without after-action-review on teaching practice in a classroom?

After-action-review is one of the perceived benefits of simulation, so it was important to investigate the impact of providing simulation with after-action-review. The following research questions and null hypotheses apply to investigating the effects of TeachLivE with an after-action-review process on teachers’ practices in their classrooms:

*Research Question 4, 5, and 6:* Are there differential effects of TeachLivE on teacher practice in a classroom based on whether or not teachers received online PD?

*Question 4:* on percentage of describe/explain questions asked during a class lesson?

*Question 5:* on percentage of specific feedback given during a class lesson?

*Question 6:* on sum score of the Teacher Practice Observation Tool from a class lesson?

**Research Questions for Student Academic Outcomes**

As noted earlier, teachers influence student learning and academic outcomes (Darling-Hammond, 2003; Hattie, 2003). With this critical variable in mind, we hypothesized that as teachers increased their frequency of strategic practices, students’ scores on academic outcomes
would increase also. Specifically, we hypothesized that students of teachers who had received four 10-minute sessions in TeachLivE would have larger academic gains than students of teachers who did not. With regard to student outcomes, we had the following research questions:

**Research Question 7:** Are there differential effects of TeachLivE on student scores based on whether or not their teachers received online PD?

**Method**

**Participant Characteristics**

Data analyzed in this study were collected during the first year of a three-year project at 10 separate research locations comprised of university and school district partners. Participants were practicing middle school mathematics teachers and were the primary teachers of record. Teachers who were not the primary teachers of record or who did not teach a middle school mathematics class (grades six, seven, or eight) were excluded from the study. No restrictions were made based on education level of teacher, number of years teaching, level of class taught, subject area within mathematics taught, or any other demographic characteristics. Overall, 135 teachers completed the study. Demographic data for all participating teachers are presented in Table 1.

**Table 1. Teacher Demographic Data**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control (n = 35)</th>
<th>PD Only (n = 35)</th>
<th>TLE Only (n = 35)</th>
<th>LE &amp; PD (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Professional licensure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>32 (91)</td>
<td>32 (91)</td>
<td>31 (89)</td>
<td>29 (97)</td>
</tr>
<tr>
<td>No</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>No response</td>
<td>3 (9)</td>
<td>2 (6)</td>
<td>4 (11)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>If licensed, is license in math?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>26 (74)</td>
<td>25 (71)</td>
<td>25 (71)</td>
<td>26 (87)</td>
</tr>
<tr>
<td>No</td>
<td>6 (17)</td>
<td>7 (20)</td>
<td>6 (17)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>No response</td>
<td>3 (9)</td>
<td>3 (9)</td>
<td>4 (11)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Area of certification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grades 5-9 math only</td>
<td>13 (37)</td>
<td>15 (43)</td>
<td>9 (26)</td>
<td>10 (33)</td>
</tr>
<tr>
<td>Grades 6-12 math only</td>
<td>6 (17)</td>
<td>8 (23)</td>
<td>12 (34)</td>
<td>11 (37)</td>
</tr>
<tr>
<td>Other</td>
<td>9 (26)</td>
<td>3 (9)</td>
<td>6 (17)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Grades 5-9 &amp; 6-12 math</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Grades 5-9 math &amp; other</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Grades 6-12 math &amp; other</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Grades 5-9 &amp; 6-12 math, &amp; other</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>No response</td>
<td>5 (14)</td>
<td>7 (20)</td>
<td>6 (17)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Grade levels taught</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-8 only</td>
<td>17 (49)</td>
<td>14 (40)</td>
<td>18 (51)</td>
<td>19 (63)</td>
</tr>
<tr>
<td>K-5 &amp; 6-8</td>
<td>5 (14)</td>
<td>7 (20)</td>
<td>6 (17)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>6-8 &amp; 9-12</td>
<td>6 (17)</td>
<td>10 (29)</td>
<td>7 (20)</td>
<td>8 (27)</td>
</tr>
<tr>
<td>K-5, 6-8, &amp; 9-12</td>
<td>4 (11)</td>
<td>2 (6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>No response</td>
<td>3 (9)</td>
<td>2 (6)</td>
<td>4 (11)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
Table 1. Teacher Demographic Data (continued)

<table>
<thead>
<tr>
<th>Variable (Cont's)</th>
<th>Control ((n = 35))</th>
<th>PD Only ((n = 35))</th>
<th>TLE Only ((n = 35))</th>
<th>TLE &amp; PD ((n = 30))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest academic level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor's</td>
<td>17 (49)</td>
<td>19 (54)</td>
<td>18 (51)</td>
<td>21 (70)</td>
</tr>
<tr>
<td>Master's</td>
<td>15 (43)</td>
<td>14 (40)</td>
<td>13 (37)</td>
<td>9 (30)</td>
</tr>
<tr>
<td>No response</td>
<td>3 (9)</td>
<td>2 (6)</td>
<td>4 (11)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Area of master's degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math education</td>
<td>2 (6)</td>
<td>4 (11)</td>
<td>4 (11)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Educational leadership</td>
<td>5 (14)</td>
<td>2 (6)</td>
<td>2 (6)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Other</td>
<td>7 (20)</td>
<td>9 (26)</td>
<td>7 (20)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>Educational leadership &amp; other</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Not applicable</td>
<td>14 (40)</td>
<td>13 (37)</td>
<td>14 (40)</td>
<td>16 (53)</td>
</tr>
<tr>
<td>No response</td>
<td>6 (17)</td>
<td>7 (20)</td>
<td>8 (23)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Years teaching math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One year</td>
<td>2 (6)</td>
<td>6 (17)</td>
<td>7 (20)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Two years</td>
<td>2 (6)</td>
<td>4 (11)</td>
<td>2 (6)</td>
<td>5 (17)</td>
</tr>
<tr>
<td>Three years</td>
<td>2 (6)</td>
<td>1 (3)</td>
<td>5 (14)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Four years</td>
<td>2 (6)</td>
<td>2 (6)</td>
<td>0 (0)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>5-10 years</td>
<td>13 (37)</td>
<td>9 (26)</td>
<td>9 (26)</td>
<td>9 (30)</td>
</tr>
<tr>
<td>More than 10 years</td>
<td>11 (31)</td>
<td>11 (31)</td>
<td>8 (23)</td>
<td>8 (27)</td>
</tr>
<tr>
<td>No response</td>
<td>3 (9)</td>
<td>2 (6)</td>
<td>4 (11)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>7 (20)</td>
<td>1 (3)</td>
<td>7 (20)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>30-39</td>
<td>11 (31)</td>
<td>7 (20)</td>
<td>10 (29)</td>
<td>9 (30)</td>
</tr>
<tr>
<td>40-49</td>
<td>6 (17)</td>
<td>12 (34)</td>
<td>6 (17)</td>
<td>11 (37)</td>
</tr>
<tr>
<td>50 or above</td>
<td>8 (23)</td>
<td>13 (37)</td>
<td>8 (23)</td>
<td>3 (10)</td>
</tr>
<tr>
<td>No response</td>
<td>3 (9)</td>
<td>2 (6)</td>
<td>4 (11)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>7 (20)</td>
<td>9 (26)</td>
<td>8 (23)</td>
<td>5 (17)</td>
</tr>
<tr>
<td>Female</td>
<td>25 (71)</td>
<td>24 (69)</td>
<td>23 (66)</td>
<td>25 (83)</td>
</tr>
<tr>
<td>No response</td>
<td>3 (9)</td>
<td>2 (6)</td>
<td>4 (11)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Asian</td>
<td>2 (6)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Black</td>
<td>2 (6)</td>
<td>3 (9)</td>
<td>3 (9)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4 (11)</td>
<td>2 (6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>White</td>
<td>24 (69)</td>
<td>26 (74)</td>
<td>28 (80)</td>
<td>27 (90)</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>No response</td>
<td>3 (9)</td>
<td>2 (6)</td>
<td>4 (11)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
Sampling Procedures

Participants were identified via a convenience sampling plan. Two hundred middle school teachers were initially recruited across 10 separate research locations. At each partnership site, teachers were self-nominated or nominated by their supervisors with the intent of receiving innovative, technology-rich PD in mathematics. Of those teachers, 157 completed a one-hour orientation and began participation in the research project, resulting in 79% participation of the sample approached. Participation was entirely voluntary with minimal to no compensation provided.

Data were collected in two settings: the teachers’ real classrooms and in the classroom simulator. Teachers were observed in middle school classrooms located in six states: Florida, Louisiana, Michigan, Mississippi, New York, and Utah. School settings ranged from urban, suburban, and rural with public or private enrollment. Classroom simulators were located at 10 client sites across the country in rooms at university or school district partner sites.

Teachers voluntarily participated in the PD activities. Incentives for participation varied at each site based on local conventions and were valued at less than $200 per teacher, but the research team leading the study did not offer direct compensation. Most incentives came in the form of a stipend or points awarded for PD supplied by the district; however, some districts did not provide financial support, and those teachers were personally motivated by professional learning. The professional learning activities described for recruiting purposes were: access to lesson plans and resources aligned to the CCS in mathematics, including the potential of working in a computer-simulated environment with innovative technology. Institutional review boards at each site and within each school district examined and approved all procedures, granting permission to conduct research.

Sample Size, Power, and Attrition

The intended sample size was 200 participants; however, many teachers who expressed initial interest did not complete orientation and provide consent for research, citing a variety of challenges. Teachers expressed concerns about limited time or reported they were not interested in receiving additional information about CCS, because their districts had professional learning initiatives related to Common Core already. Multiple districts were approached, but chose not to participate because of concerns that the professional learning would duplicate their own activities or conflict with their district initiatives. District reticence to participate was an unexpected challenge, as many researchers located at sites across the country involved in the project anticipated district support. Universities who reported strong ties to school districts at multiple levels of administration had the strongest numbers in terms of recruiting teachers who completed orientation and consented to participate.

Power analysis for sample size. In a review of literature, no similar studies were identified using a large group design for practicing teachers’ PD in a classroom simulator to offer an estimate of the effect size, so an a priori power analysis was conducted (Cohen, 1988). Power analysis for an F-test Analysis of Variance (ANOVA) within-between interactions resulted in a total sample size of 48 participants to have 80% power for detecting a medium sized effect (0.25) when employing a 0.10 criterion of statistical significance. A 0.10 criterion was selected due to
the cutting edge research’s low risk to humans; therefore, a larger Type Two error was acceptable in considering the overall findings. The projected number of participants was 200, based on funding allocated for the research project. The anticipated number of participants exceeded the suggested number of 48 participants for a medium sized power effect.

**Measures and Covariates**

Data were collected on a variety of measures from teachers and their students, including qualitative and quantitative measures. See Table 2 for an overview of data sources.

**Table 2. Overview of Data Sources**

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Individuals</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher demographics</td>
<td>Teachers</td>
<td>Teacher Participant Orientation Questionnaire</td>
</tr>
<tr>
<td>Teaching practice in TeachLivE classroom simulator</td>
<td>Teachers</td>
<td>TeachLivE After-Action-Review System (TeachLivEAARS)</td>
</tr>
<tr>
<td>Teacher perceptions of TeachLivE experience</td>
<td>Teachers</td>
<td>TeachLivE Presence Questionnaire</td>
</tr>
<tr>
<td>Teacher perceptions of preparation after TeachLivE</td>
<td>Teachers</td>
<td>TeachLivE Perceptions Questionnaire</td>
</tr>
<tr>
<td>Teaching practice in classroom</td>
<td>Teachers</td>
<td>Teacher Practice Observation Tool</td>
</tr>
<tr>
<td>Student academic performance</td>
<td>Students</td>
<td>Curriculum-based measure</td>
</tr>
<tr>
<td>Student demographics</td>
<td>Students</td>
<td>Cross-reference Demographic Sheet</td>
</tr>
</tbody>
</table>

**Teacher data.** All teachers were observed teaching in their classrooms pre- and post-treatment using quantitative and qualitative observations on the Teacher Practice Observation Tool (TPOT, see Appendix A). Teachers responded to demographic questions during orientation. During classroom observations, data were collected on the frequency of HLPs determined to increase the likelihood that these teaching behaviors would have a positive effect on students’ learning outcomes (Teaching Works, 2014), modified sub-constructs from the 2011 Danielson Framework for Teaching Evaluation Instrument (Danielson, 2011), and observations of practice using qualitative field notes. Data were collected in five-minute intervals, rotating across constructs (see Appendix A), so observers focused on one construct at a time during the class. For the teachers who experienced the classroom simulator, data also were collected on their sense of presence and preparedness after the four sessions of virtual rehearsal, and the frequency of HLPs exhibited in each session in the simulated environment. Within the classroom simulator, the observers did not divide their observations into intervals, and instead focused simultaneously on two variables: frequency of questioning and feedback throughout the 10-minute session. The researchers gathered the information provided in Table 3 as to the variables observed in the real classroom and the classroom simulator.
Table 3. Variables Observed in the Classroom and Classroom Simulator.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Classroom</th>
<th>Classroom Simulator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One period</td>
<td></td>
</tr>
<tr>
<td></td>
<td>45 –95 minutes</td>
<td>10-minute session</td>
</tr>
<tr>
<td></td>
<td>(2 observations)</td>
<td>(4 observations)</td>
</tr>
<tr>
<td><strong>High-leverage Practices</strong></td>
<td>Questioning</td>
<td>Questioning</td>
</tr>
<tr>
<td></td>
<td>Feedback</td>
<td>Feedback</td>
</tr>
<tr>
<td></td>
<td>Wait Time</td>
<td></td>
</tr>
<tr>
<td><strong>Type of Data Collected</strong></td>
<td>Frequency per 1 minute interval</td>
<td>Frequency per 10-minute session</td>
</tr>
<tr>
<td></td>
<td>every 5 minutes</td>
<td>every 5 minutes</td>
</tr>
</tbody>
</table>

**Sub-constructs from 2011**

8 sub-constructs

**Danielson Framework for Teaching**

- Sum of observer ratings at the end of the observation
- Structured protocol of field notes for 2 minutes every 5 minutes

**High-leverage practices.** Using research from the Measures of Effective Teaching project and definitions developed from that research and the field, the following data were collected. The HLP behaviors observed were the teachers’ frequency and type of eliciting and interpreting individual students’ thinking (HLP #3). Specifically, data were collected on frequency of:

- describe/explain questions: any content question that requests a description or explanation of a mathematical object, non-prescribed solution methods, or a reason why something is true or not true;

- short response questions: any content question that requests a relatively short response, such as vocabulary, numbers, formulas, single rules, prescribed solution methods, or an
answer for computation; requests that a student read the response from a notebook or textbook; and requests that a student choose among alternatives;

- yes/no questions: any content question that elicits a simple “yes” or “no” response.

In the classroom simulator, frequency data were collected. Each session lasted 10 minutes so the frequency and type of instances for each behavior were noted. In the teachers’ classroom, lessons varied in length (45 to 95 minutes), so a percentage of describe/explain questions was calculated. To calculate percentage, the occurrences of describe/explain questions were divided by the sum of all questions (describe/explain, short, and yes/no) and multiplied by 100 to arrive at a percentage score for the observation.

Frequency data also catalogued the type of feedback teachers gave students. Effective feedback is specific, not overwhelming in scope, focused on the academic task, and supports students’ perceptions of their own capability (HLP #12). The teachers’ type of feedback exhibited in the simulator was separated into two categories and defined as:

- specific feedback: teacher’s verbal response to a student’s statement/question that meets all of the following criteria: positive in tone, exact, focused on the academic tasks, and supporting students’ perception of their capability or reinforcing student’s capabilities;

- general feedback: teacher’s verbal response to a student’s statement/question that includes at least one of the following: critical, negative in tone, one word responses, ambiguous or vague, lacks relation to academic tasks.

As with describe/explain questions, in the classroom simulator, frequency data of specific feedback were collected. In the teachers’ real classrooms, the time in each class period varied, so a percentage of specific feedback was calculated. To calculate percentage, the occurrences of specific feedback were divided by the sum of all feedback (specific plus general) and multiplied by 100 to arrive at a percentage score for the observation.

Finally, frequency data were collected on the amount of time teachers waited after asking questions as a means of providing students with sufficient time to think about their response, to reflect on the comments of their classmates, and to deepen their understanding (HLP #3). Brophy and Good (1986) recommended three to five seconds of wait time after a question is posed. For the purposes of this study, wait time was defined as a dichotomous variable, separating it into time greater than or equal to three seconds or time less than three seconds:

- wait time less than three seconds: number of seconds beginning exactly after the termination of a teacher question and ending when the teacher repeats, rephrases, answers, or asks a new question is less than three seconds;

- wait time greater than or equal to three seconds: number of seconds beginning exactly after the termination of a teacher question and ending when the teacher repeats, rephrases, answers, or asks a new question is greater than or equal to three seconds.
To calculate percentage of wait time, first a dichotomous variable was created, categorizing wait time as less than three seconds (WT<3) or greater than or equal to three seconds (WT≥3). Then, the occurrences of WT≥3 were divided by the sum of WT≥3 and WT<3 and multiplied by 100 to arrive at a percentage score for the observation.

**Sub-constructs from 2011 Danielson Framework for Teaching.** Eight sub-constructs correlated with student achievement were identified from the 2011 Danielson Framework for Teaching Evaluation Instrument (Measures of Effective Teaching Project, 2010). Key words from Danielson’s indicators were chosen to create an abbreviated version to be used in classroom observations combined with the collection of frequency data in relation to describe/explain questions, specific feedback, and wait time. Danielson’s four levels of performance (i.e., unsatisfactory, basic, proficient, distinguished) were the basis for a four-point scale for each sub-construct: establishing a culture for learning, engaging students in learning, managing student behavior, managing classroom procedures, communicating with students, using questioning and discussion techniques, creating an environment of respect and rapport, and using assessment in instruction. Further, qualitative data were collected during the classroom observation on each sub-construct listed above using a field notes method. See Appendix A for the TPOT that includes each sub-construct and associated scale.

**TPOT Development.** For classroom observations, all data were gathered on a single form, the TPOT. The class period was divided into five-minute overarching intervals, which were subdivided into smaller intervals. Observers focused on one aspect of the HLPs during each interval, as follows: minute one: frequency of questions (describe/explain, short answer, yes/no); minute two: frequency of wait time (three seconds or more, less than three seconds); minute three: frequency of feedback (general or specific). Space was provided for comments within each interval. The final two minutes of the five-minute interval were dedicated to observation, field notes, and scoring of the Danielson sub-constructs. At the close of the lesson, observers completed a final summary of the observation, assigning a final score on each Danielson sub-construct and providing a written rationale. Data collectors observed approximately 30% of teachers for inter-observer agreement; at the end of each observation, observers discussed divergent scores and documented any changes with a justification (see TPOT in Appendix A).

Prior to commencing the study, the TPOT was piloted with a purposive sample of four certified secondary teachers, two teacher preparation college faculty members, a national research expert on teacher pedagogy, and a content area expert. The research team consulted with all members during the development and validation of classroom-based observational techniques in order to ensure that domains on the observation tool were in line with secondary classroom practices and adhered to evidence-based instructional practices. Several meetings were held to discuss the initial target teaching practices, refine operational definitions, and explore options for coding. Two separate institutions for teacher professional learning provided observational domains, and representatives provided an expert validation of the domains included in the TPOT. The research team also observed a variety of teaching videos and arrived at 100% agreement on each high leverage practice and the sub-constructs from the Danielson Framework. This process provided a basis for validating that the observational tool and protocols were clear and meaningful. Reliability estimates related to each variable are provided in the results section.
**TeachLivE After-Action-Review System (TeachLivEAARS).** During each TeachLivE session, the teachers’ virtual rehearsal was transmitted via secure Skype video and audio connection. The transmissions were recorded and coded for pedagogical strategy analysis using TeachLivEAARS software. TeachLivEAARS is a video tagging software integrated with the TeachLivE classroom simulator that records sessions, compresses the video to a smaller format, and then sends the video over a secure network to be stored at the originating research site computer containing the TeachLivE software. During each session, videos were tagged for frequency and type of questions and feedback. A beta version of TeachLivEAARS was used in year one of the project, and brought about intermittent issues with recording and exporting of tags, so data also were collected using a paper and pencil backup to maintain integrity.

**TeachLivE questionnaires.** Each teacher that entered the classroom simulator was administered two researcher-created questionnaires:

- **TeachLivE Presence Questionnaire (Hayes, Hardin, & Hughes, 2013):** Teachers responded to questions about their simulation experience related to suspension of disbelief, presence, fidelity, and immersion.
- **TeachLivE Perceptions Questionnaire (Hardin, Hayes, & Hughes, 2013):** Teachers also responded to items about how virtual rehearsal in the classroom simulator prepared them for teaching in their own classrooms.

**Student data.** Data were also collected from middle school students in the participating teachers’ classrooms on student performance on a curriculum-based measure of algebraic equations based on the NAEP. Ten items from the eighth grade 2011 NAEP were used to collect information about student achievement. Teachers were instructed to give students 20 minutes to complete the assessment. Teachers also provided general information about student demographics.

**Methods used to enhance the quality of measurements.** Due to the national nature of the study, researchers and observers were at sites across the country, and this presented challenges for observational teams in terms of training and reliability of observations. Therefore, all data collectors were trained online using a combination of asynchronous assessment and synchronous data collection training on the constructs (e.g., Danielson sub-constructs and HLPs) and methods (e.g., frequency counts during rotating intervals as described above) for data collection. Data collectors used the asynchronous online modules to demonstrate proficiency with the content of observations. Each practice was defined and a case example was provided. Observers had to pass a multiple-choice content assessment with 90% accuracy for the asynchronous portion of the training. The synchronous online training consisted of a series of activities delivered via a video conferencing platform that exposed observers to operational definitions and required the collection of frequency counts in real time while watching a video online as a group to simulate classroom observations. Each observer was checked for reliability during the online training and required to complete a synchronous online activity with 90% accuracy.
Research Design

The research design was a group randomized trial, consisting of four groups of teachers measured pre-post in the classroom, and two of the groups were measured four times in the classroom simulator. To prevent treatment diffusion across conditions, teachers at each school were grouped together in a unit and randomly assigned to one of four treatment conditions. The random assignment procedure took place at all 10 partnership sites, resulting in four experimental groups.

Interventions

Teachers received varying levels of PD based on a lesson plan aligned to the Common Core standards, Classroom Challenges: Solving Linear Equations in One Variable (Mathematics Assessment Resource Service, 2012) and were assigned to one of four groups: Group 1 teachers served as a comparison group and received lesson plans only; Group 2 teachers received lessons and online PD; Group 3 teachers received lessons and TeachLivE; and Group 4 teachers received lessons, online PD, and TeachLivE. See Figure 1 for an overview of the four treatment groups.

Figure 1. Overview of Four Treatment Groups

**Group 1 (G1): Comparison.** As with all four groups, teachers in the comparison group received the mathematics lesson plan on linear equations via email. They were encouraged to explore the lesson plan and implement it with their students during the course of the school year. They were given no other intervention as a course of this study, but did receive any PD provided by their district throughout the course of the school year.

**Group 2 (G2): Online PD.** Teachers in G2 received a digital copy of the lesson plan (like the teachers in G1), as well as one 40-minute session of online PD with a nationally recognized, doctoral-level expert in the Classroom Challenges curriculum, delivered via the Adobe Connect platform. The platform required a Macintosh or PC computer with a web camera, microphone, and speakers. The Adobe Connect program was loaded onto the computer’s web browser to create an online learning environment (Adobe Systems Incorporated [ASI], 2014). Adobe Connect provided a dedicated online meeting room at a secure web address that
did not require a software download. Teachers could see and hear the PD provider via a video-conferencing pod and could respond in real-time using their microphones or a chat box feature; however, teachers’ web cameras were not enabled due to the short length of the training in comparison with the technical requirements needed to enable their video. Instead, teachers communicated via their microphones or, in a few cases, a chat box window if the microphone was not in operation. The online PD occurred on six separate occasions to accommodate teachers with approximately 10 participants per session, and the same curriculum and format was used each time. After a 10-minute orientation to Adobe Connect, the online PD content included a discussion of the five strategies of formative assessment: (a) clarifying and sharing learning intentions and criteria for success; (b) engineering effective discussion, questions, activities, and tasks that elicit evidence of learning; (c) providing feedback that moves the learner forward; (d) activating students as instructional resources for each other; and (e) activating students as owners of their own learning (Thompson & Wiliam, 2008). After the conclusion of the discussion, teachers took part in an analysis of five authentic student work samples in response to a formative assessment included in the lesson (see Appendix B for student work samples used in the groups going into the simulator), followed by another discussion about questioning strategies and feedback for students. Teachers were asked to create questions and provide feedback for students based on the provided examples of student work. The treatment length of online PD was 10 minutes of orientation followed immediately by 40 minutes of PD; this set amount of time equaled the amount of time spent in the simulator.

**Group 3 (G3): TeachLivE.** Teachers in G3 received a digital copy of the lesson (like teachers in G1 and G2), as well as four 10-minute virtual rehearsal sessions in the TeachLivE classroom simulator. In the simulator, teachers attended individually and interfaced with a computer-generated, animated student population of five middle school avatars puppeteered by the interactor. The software is programmed to react to certain commands of the teacher and the interactor, with the purpose of increasing the teacher’s aptitude in the classroom. Classroom simulators at 10 client sites across the country were connected via secure server to the Synthetic Reality Laboratory at the University of Central Florida, which served as the central distribution point for TeachLivE and provided fidelity of treatment that all sessions were controlled at the primary research site. For operation at the teacher client sites, the simulator required a computer with TeachLivE software, large display monitor, webcam, lavaliere microphone, speakers, system for tracking movement, and an Internet connection. A session facilitator, trained on how to use the software and enact the research procedures, facilitated the sessions and collected the data. At the server where the interactor was located, a computer with TeachLivE software, monitor, and motion tracking devices were needed to operate the system. The teachers experienced computer-simulated classroom activities with the student-avatars as they would with human students in a traditional classroom. Visits to the simulator took place over the course of four to six weeks.

As with G2 during the online PD, teachers in G3 participated in 10-minute sessions to orient them to the TeachLivE system. Data were not collected during the orientation session, as users were not teaching content but interacting with the student-avatars with the objective of learning about their class. After orientation, teachers received four 10-minute sessions in TeachLivE (sessions 1 through 4) to take part in virtual rehearsal (i.e., targeted practicing of a skill in a virtual environment), with data on targeted behaviors gathered during each session.
Prior to the virtual rehearsal, teachers were given the same student work samples used in the online PD (see Appendix B), but in this condition, teachers were told that each work sample was a product of a specific student-avatar. Teachers were instructed to rehearse the whole class discussion portion of a specified Classroom Challenges lesson (Solving Linear Equations in One Variable; Mathematics Assessment Resource Service, 2012) and, at the close of each session, they took part in an after-action-review of their performance. After-action-review consisted of three parts: (a) teachers were asked to estimate their frequency of higher order questions and specific feedback, (b) teachers were shown their actual frequency of observed behaviors in the session, and (c) teachers were asked how they intended to use this information. Upon completion of the after-action-review, teachers returned to the simulation for another session. After orientation, teachers typically took part in two 10-minute sessions and returned within a month for another two 10-minute sessions.

**Group 4 (G4): Online PD combined with TeachLivE.** Teachers in G4, the TeachLivE and Online PD combined condition: (a) received the lesson plan, (b) participated in the online PD, and (c) engaged in virtual rehearsal in TeachLivE as outlined above. Teachers did not enter TeachLivE until they had completed the online PD.

**Results**

At the beginning of the research study, 157 teachers completed orientation and were grouped by school, then randomly assigned to four groups in a randomized group design nested within school (see Table 4); however, teacher requests to change treatment groups necessitated a modification resulting in the final quasi-experiment design (Step 1 of participant flow through the quasi-experiment). Many teachers in the study were concerned with scheduling, because the PD sessions occurred after school hours. As a result, once scheduling began, eleven teachers requested to be removed from the study or to be changed to a different PD treatment group (Step 2 of participant flow). Three teachers requested to be moved from G1 to G3 to increase their level of treatment (from lesson plans to lesson plans plus TeachLivE) to potentially receive benefits of PD. Six teachers wanted to continue participation, but could not complete the PD activities due to prior commitments, so they were moved from G2, G3, or G4, which required outside-class PD activities, to G1, which required only in-class activities (i.e., being observed teaching to their real class). Two teachers moved from G4 (the combined online PD and TeachLivE group); one moved to G2 and received only online PD, while the other moved to G2 and received only TeachLivE. While changes in treatment group did violate random assignment procedures, all changes occurred prior to interventions, so that no teachers received a partial intervention and then switched to another group midway through an intervention. Teachers attended events individually (e.g., teachers had a selection of the online PD and TeachLivE sessions to choose from); therefore group assignment could occur prior to the intervention. The most common reason cited for attrition/change in assignment was teacher stress level. Anecdotal evidence from the research sites indicated that some teachers reported feeling “overwhelmed” with teaching duties and were concerned about an additional responsibility; however, data were not collected on teachers’ reasons for attrition. Table 4 outlines the participant flow through each stage of the study. First, teachers were randomly assigned to treatment groups; then teachers made requests to be changed from their group and some teachers were lost to attrition (Step 3 of participant flow), resulting in the final number of teachers per treatment group.
Table 4. Participant Flow through the Quasi-Experiment

<table>
<thead>
<tr>
<th>Step 1: Teachers oriented and schools randomly assigned</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>34</td>
<td>41</td>
<td>40</td>
<td>42</td>
<td>157</td>
</tr>
</tbody>
</table>

Step 2: Changes in groups per teacher request

| From G1 to G3 | -3 | 3 |
| From G2 to G1 | 3  | -3|
| From G3 to G1 | 2  | -2|
| From G4 to G1 | 1  | -1|
| From G4 to G3 | 1  | -1|
| From G4 to G2 | 1  | -1|

Step 3: Teachers lost to attrition

| -2 | -4 | -7 | -9 | -22 |

Step 4: Final number of teachers per group

| 35 | 35 | 35 | 30 | 135 |

Treatment Fidelity

Fidelity checks were in place throughout the study. All teachers received the lesson plan in digital format, as evidenced by a checklist of teacher contact information at each site. A facilitator monitored online PD and checked for fidelity of implementation at each phase of the online session. All online PD sessions were delivered at 100% accuracy as evidenced by a lesson plan checklist outlining the content. During the TeachLivE sessions, the facilitator followed a detailed procedural checklist to turn on and operate the software for the simulation, ensuring fidelity of implementation.

Data Analysis

Teaching practices were defined on three distinct dimensions pre- and post-intervention: (a) describe/explain questions (DE), (b) specific feedback (SF), and (c) summary score on the TPOT (TPOT Sum). Maxwell’s (2001) recommendation of moderate correlation (0.3 – 0.7) was used as a threshold for all variables to determine if it was appropriate to conduct a multivariate analysis of variance. Wait time was excluded from the analysis because the researchers predicted no significant findings. In the case of the variables under investigation, the majority did not meet correlation thresholds, so analysis of variance (ANOVA) tests were more appropriate. See Table 5 for correlations of dependent variables.
Table 5. Correlations of Dependent Variables.

<table>
<thead>
<tr>
<th></th>
<th>DE Pre</th>
<th>DE Post</th>
<th>SF Pre</th>
<th>SF Post</th>
<th>TPOT Sum Pre</th>
<th>TPOT Sum Post</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DE Pre</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DE Post</td>
<td>.130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF Pre</td>
<td>.292**</td>
<td>-0.028</td>
<td>1</td>
<td></td>
<td>.076</td>
<td></td>
</tr>
<tr>
<td>SF Post</td>
<td>.208*</td>
<td>.259**</td>
<td>.076</td>
<td>1</td>
<td>.118</td>
<td>.199</td>
</tr>
<tr>
<td>TPOT Sum Pre</td>
<td>.371**</td>
<td>.185*</td>
<td>.183*</td>
<td>.118</td>
<td>1</td>
<td>.633**</td>
</tr>
<tr>
<td>TPOT Sum Post</td>
<td>.220</td>
<td>.353**</td>
<td>.171</td>
<td>.199</td>
<td>.633**</td>
<td>1</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Specific statistical tests used and variables under consideration are described in detail in the results section. The results are divided by classroom simulator data and classroom data, and then further subdivided by research question.

Classroom Simulator Results

Teachers in G3 and G4, who received PD in TeachLivE, were administered a questionnaire gathering information about their perceptions of presence in the simulator. Over 80% of teachers agreed that the TeachLivE virtual classroom felt like a real classroom and over 90% of teachers agreed that the virtual students accurately represented the kinds of people that existed in the real world.

Research Questions 1 and 2: Differences in performance over time with simulation and differential effects. To examine performance of teachers over four 10-minute sessions, a two-factor mixed design ANOVA was performed. Time (four sessions) was cast as a within-subjects factor, and condition (two levels, online PD and no online PD) functioned as a between-subjects factor, with dependent variables of DE for question 1 and SF for question 2. Due to the novel nature of the intervention (e.g., dearth of group design research identified on simulation in teacher education), an alpha level of .10 was established to judge statistical significance. Partial eta squared was used to interpret effect size rather than eta squared because a multifactor design was used (Pierce, Block, & Aguinis, 2004) and we wanted to be able to compare effects across different factorial designs used in the study (Levine & Hullet, 2002).
**Question 1: DE teacher practice.** Two observers collected data on frequency of DE questions asked by teachers per TeachLivE session. Pearson’s correlation provided a basis for interpreting reliability of scores between observers during each session (session 1, \( r = .952 \); session 2, \( r = .820 \); session 3, \( r = .660 \); session 4, \( r = .986 \)). Results from a two-factor mixed design ANOVA indicated no differential effects for teachers who did or did not get online PD \( (F(3,171) = .735, p = .532, \eta^2_p = .13) \). However, a significant time effect was identified \( (F(3,171) = 9.993, p = .000, \eta^2_p = .149) \). Pallant (2007) recommends interpreting partial eta squared using Cohen’s (1988) guidelines for eta squared effect size: small (.01), medium (.06), or large (.14). Mean scores increased at each session (see Table 6).

Table 6. Mean DE Questions across 10-minute TeachLivE Sessions

<table>
<thead>
<tr>
<th>PD Factor</th>
<th>TeachLivE Sessions</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Session 1</td>
<td>Session 2</td>
<td>Session 3</td>
<td>Session 4</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>No Online PD</td>
<td>n=34</td>
<td>5.1 (4.2)</td>
<td>7.6 (4.9)</td>
<td>8.4 (5.3)</td>
</tr>
<tr>
<td>Online PD</td>
<td>n=25</td>
<td>6.5 (4.2)</td>
<td>7.9 (4.5)</td>
<td>7.9 (5.9)</td>
</tr>
<tr>
<td>Total</td>
<td>n=59</td>
<td>5.7 (4.2)</td>
<td>7.8 (4.7)</td>
<td>8.2 (5.5)</td>
</tr>
</tbody>
</table>

**Question 2: SF teacher practice.** Two observers collected data on frequency of SF given by teachers to student-avatars per TeachLivE session. Reliability of scores between observers during each session was calculated (session 1, \( r = .928 \); session 2, \( r = .872 \); session 3, \( r = .811 \); session 4, \( r = .790 \)). Results from a two-factor mixed design ANOVA indicated no differential effects for teachers who did or did not get online PD \( (F(3,168) = 1.989, p = .118, \eta^2_p = .034) \). Yet, a significant time effect was found \( (F(3,168) = 2.306, p = .079, \eta^2_p = .040) \). Again, mean scores increased at each session (see Table 7).

Table 7. Mean SF Questions across 10-minute TeachLivE Sessions

<table>
<thead>
<tr>
<th>PD Factor</th>
<th>TeachLivE Sessions</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Session 1</td>
<td>Session 2</td>
<td>Session 3</td>
<td>Session 4</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>No Online PD</td>
<td>n=34</td>
<td>6.2 (5.1)</td>
<td>8.3 (6.0)</td>
<td>8.7 (4.8)</td>
</tr>
<tr>
<td>Online PD</td>
<td>n=24</td>
<td>6.9 (4.5)</td>
<td>6.7 (3.6)</td>
<td>6.3 (3.9)</td>
</tr>
<tr>
<td>Total</td>
<td>n=58</td>
<td>6.5 (4.8)</td>
<td>7.8 (5.1)</td>
<td>7.7 (4.6)</td>
</tr>
</tbody>
</table>

Figure 2 shows the trend of mean scores of frequency of instances of DE and SF across sessions.
Classroom Results

To investigate the effects on teacher practice in a classroom setting, teacher behavior was considered in TeachLivE with and without an integrated after-action-review process. Two observers collected data during classroom observations pre- and post-treatment. A description of the statistical analyses and variables under investigation follows.

Simulation without after-action-review. To examine performance of teachers in a classroom after TeachLivE sessions without after-action-review, we performed a three-factor mixed design ANOVA with between-subjects factors of simulation (TeachLivE and no TeachLivE) and online PD (online PD and no online PD), and a within-subjects factor of time (pre- and post-intervention). The dependent variable was percentage of wait time three seconds or more (WT>3). Due to the novel nature of the intervention, an alpha level of .10 was established to judge statistical significance.

Question 3: Wait Time (WT) teacher practice. An observer collected data on frequency of WT>3 in a class, and two observers observed 30% of classes to establish inter-rater reliability. Reliability of scores between observers during both observations was calculated (pre-intervention, $r = .338$; post-intervention, $r = .718$). Data should be interpreted with caution, due to low reliability scores between classroom observers during the first observation. Scores were not normally distributed, as assessed by Shapiro-Wilk’s test ($p < .05$); however, ANOVAs are considered to be robust to deviations from normality. There was homogeneity of variances for frequency of wait time at both pre ($p = .827$) and post-intervention ($p = .161$), as assessed by Levene’s test for equality of variances.

Results from a three-factor mixed design ANOVA indicated a nonsignificant effect for the three-way interaction effects of time, simulation, and online PD ($F(1,130) = 1.003, p = .318$, $\eta^2_p = .008$). No effects were found for simple two-way interaction between time and simulation.
online PD decreased their questions by 3%, whereas those who did not in the three intervention (not a statistically significant difference between groups assigned to TeachLivE). When comparing the effects of TeachLivE over time, there was no interaction between time and online PD ($F(1,130) = 1.580, p = .211, \eta^2_p = .012$).

**Simulation with after-action-review.** TeachLivE simulation with after-action-review may contribute to changes in teacher practice, but that effect might differ across teachers who received online PD as well. Again, we used a three-factor mixed design ANOVA to evaluate the effectiveness of TeachLivE with after-action-review. Dependent variables of DE questions, SF, and TPOT Sum were analyzed.

**Question 4: DE teacher practice.** In the teachers’ classrooms, lessons varied in length (45 to 95 minutes), so a percentage of DE questions was calculated and used as the pre-post measure. Observer reliability was evaluated using Pearson’s correlation (pre-intervention, $r = .701$; post-intervention, $r = .795$).

A three-way mixed ANOVA was conducted to understand the effects of TeachLivE, online PD, and time on percentage of DE questions asked during a lesson. Scores were not normally distributed, as assessed by Shapiro-Wilk’s test ($p < .05$); however, ANOVAs are considered to be robust to deviations from normality. There was homogeneity of variances for percentage of DE asked at both pre- ($p = .065$) and post-intervention ($p = .335$), as assessed by Levene’s test for equality of variances. Results of the three-factor mixed design ANOVA indicated no differential effect of time for online PD when combined with TeachLivE ($F(1,130) = .168, p = .682, \eta^2_p = .001$). The interaction between TLE and online was not statistically significant ($F(1,130) = .015, p = .902, \eta^2_p = .000$). There was a statistically significant two-way interaction between time and online PD ($F(1,130) = 5.735, p = .018, \eta^2_p = .042$) and time and TeachLivE ($F(1,130) = 3.479, p = .064, \eta^2_p = .026$). Statistical significance of a simple main effect was accepted at a Bonferroni-adjusted alpha level of .050. All pairwise comparisons were performed for statistically significant simple main effects. Bonferroni corrections were made with comparisons within each simple main effect considered a family of comparisons. Adjusted $p$-values are reported. Statistically significant differences existed at pre-intervention for those assigned to online PD, $(F(1,130) = 4.854, p = .029, \eta^2_p = .036)$, but not at post-intervention $(F(1,130) = 1.204, p = .902, \eta^2_p = .275)$, which suggests a difference in groups at pre-intervention. For those assigned to the online PD groups, mean percentage DE was higher at pre-intervention than for those who were not, with a mean difference of 5.7% (90% CI, 0.014 to 0.100), $p = .29$. However the overall focus of the research was TeachLivE with online PD only as a secondary consideration. When comparing the effects of TeachLivE over time, there was not a statistically significant difference between groups assigned to TeachLivE at pre-intervention $(F(1,130) = 1.274, p = .261, \eta^2_p = .010)$, but there was a post-intervention $(F(1,130) = 9.827, p = .002, \eta^2_p = .070)$, suggesting significant effects for TeachLivE as an intervention. Mean percentage DE was higher at post-intervention for those who received TeachLivE than those who did not, with a mean difference of 10% (90% CI, 0.048 to 0.154), $p = .002$. Because the three-way interaction was not significant, it is appropriate to compare performance of teachers pre-to post-intervention on both TeachLivE and online PD. Teachers who received the online PD decreased their questions by 3%, whereas those who did not increased questions by...
7%; however, significant differences between groups pre-intervention existed. Conversely, TeachLivE teachers increased DE questions by 6%, whereas teachers who did not get TeachLivE decreased them by 2%, and no significant differences existed pre-intervention. See Table 8 for mean changes from pre to post.

Table 8. Means Changes in Percent DE.

<table>
<thead>
<tr>
<th>Time Factor</th>
<th>TeachLivE Factor</th>
<th>Online PD Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Pre</td>
<td>18 (17)</td>
<td>16 (14)</td>
</tr>
<tr>
<td>Post</td>
<td>24 (20)</td>
<td>14 (16)</td>
</tr>
<tr>
<td>Change</td>
<td>+6 -2</td>
<td>-3 +7</td>
</tr>
</tbody>
</table>

An a priori hypothesis was established to determine whether or not there would be differences in percentage of DE questions for teachers who received TeachLivE as compared to teachers who did not. The researchers, using a test of contrast, suggested evidence against the null hypothesis of no difference. Teachers who received TeachLivE, on average, asked a significantly higher ($t(132) = 3.198, p = .002$) percentage of DE questions at post-test ($M = 24\%$) than those who did not ($M = 14\%$).

**Question 5: SF teacher practice.** Pearson’s correlation provided a basis for interpreting reliability of scores between observers (pre-intervention, $r = .347$; post-intervention, $r = .562$). Data should be interpreted with caution, due to low reliability scores. A three-way mixed ANOVA was conducted to understand the effects of TeachLivE, online PD, and time on percentage of SF given during a lesson. Scores were not normally distributed, as assessed by Shapiro-Wilk’s test ($p < .05$); however, ANOVAs are considered to be robust to deviations from normality. There was homogeneity of variances for percentage of SF at both pre- ($p = .794$) and post-intervention ($p = .731$), as assessed by Levene’s test for equality of variances. Results of the three-factor mixed design ANOVA indicated a differential effect of time for online PD when combined with TeachLivE ($F(1,130) = 3.486, p = .064, \eta^2_p = .026$). Statistical significance of a simple two-way interaction was accepted at a Bonferroni-adjusted alpha level of .050. There was a statistically significant simple two-way interaction of TeachLivE and online PD at pre-intervention ($F(1,131) = 3.638, p = .059, \eta^2_p = .027$), but not at post-intervention ($F(1,130) = .527, p = .469, \eta^2_p = .004$). Statistical significance of a simple main effect was accepted at a Bonferroni-adjusted alpha level of .050. All pairwise comparisons were performed for statistically significant simple main effects. Bonferroni corrections were made with comparisons within each simple main effect considered a family of comparisons. Adjusted $p$-values are reported. Data are mean ± standard deviations unless otherwise stated. At pre-interventions, there was not a statistically significant difference between G1 and G2, those without TeachLivE ($F(1,131) = .321, p = .572, \eta^2_p = .002$). However, for groups with TeachLivE as a factor, G3 and G4, there was a statistically significant difference at pre-intervention ($F(1,131) = 4.411, p = .038, \eta^2_p = .033$). Mean SF scores at pre-intervention were significantly higher for those teachers in G4 (37.7% ± 27.5%) than those in G3 (24.4% ± 23.3%), a mean difference of 13.3% (90% CI, 2.8% to 23.7%), $p = .038$. Teachers in G3 had the highest gains.
(+18%) of the four treatment groups, yet their colleagues in G4, who received both TeachLivE and online PD decreased in SF (-2%), the only decrease across all four groups. See Table 9 for changes in mean percentages over time and data displayed visually in Figure 3.

Table 9. Changes in Mean Percentages of SF over Time.

<table>
<thead>
<tr>
<th>Treatment Groups</th>
<th>n</th>
<th>Pre M (SD)</th>
<th>Post M (SD)</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>35</td>
<td>29 (24)</td>
<td>34 (27)</td>
<td>+5</td>
</tr>
<tr>
<td>Online PD</td>
<td>34</td>
<td>27 (26)</td>
<td>34 (28)</td>
<td>+7</td>
</tr>
<tr>
<td>TeachLivE</td>
<td>35</td>
<td>24 (23)</td>
<td>42 (27)</td>
<td>+18</td>
</tr>
<tr>
<td>TeachLivE &amp; Online PD</td>
<td>30</td>
<td>38 (27)</td>
<td>36 (24)</td>
<td>-2</td>
</tr>
</tbody>
</table>

Figure 3. Percent SF over Time.

**Question 6: TPOT sum teacher practice.** Observer reliability was evaluated using Pearson’s correlation (pre-intervention, r = .933; post-intervention, r = .970). A three-way mixed ANOVA was conducted to understand the effects of TeachLivE, online PD, and time on TPOT sum score on a lesson. Scores were not normally distributed, as assessed by Shapiro-Wilk’s test (p < .05); however, ANOVAs are considered to be robust to deviations from normality. There was a homogeneity of variances for TPOT sum at both pre- (p = .218) and post-intervention (p = .519), as assessed by Levene’s test for equality of variances. Results of the three-factor mixed design ANOVA indicated a differential effect for time for online PD when combined with TeachLivE (F(1,117) = 3.003, p = .086, η²p = .025). Statistical significance of a simple two-way interaction was accepted at a Bonferroni-adjusted alpha level of .050. There was neither a statistically significant simple two-way interaction of online PD and TeachLivE at pre-intervention (F(1, 125) = 1.180, p = .280, η²p = .009), nor post-intervention (F(1,121) = .008, p = .928, η²p = .000). As with SF, teachers who received TeachLivE without online PD had the
highest gains (+1.03) of the four treatment groups; yet their colleagues who received both TeachLivE and the online PD decreased by the largest amount (-.78). See Table 10 for changes in scores over time and data displayed visually in Figure 4.

Table 10. Changes in Mean Score of TPOT Sum over Time.

<table>
<thead>
<tr>
<th>Treatment Groups</th>
<th>n</th>
<th>M (SD)</th>
<th>M (SD)</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td></td>
<td>Post</td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>32</td>
<td>22.06 (3.75)</td>
<td>22.00 (4.20)</td>
<td>- .06</td>
</tr>
<tr>
<td>Online PD</td>
<td>32</td>
<td>21.33 (5.35)</td>
<td>21.83 (4.81)</td>
<td>+ .50</td>
</tr>
<tr>
<td>TeachLivE</td>
<td>32</td>
<td>21.63 (4.53)</td>
<td>22.66 (3.97)</td>
<td>+1.03</td>
</tr>
<tr>
<td>TeachLivE &amp; Online PD</td>
<td>27</td>
<td>23.19 (3.88)</td>
<td>22.41 (4.49)</td>
<td>- .78</td>
</tr>
</tbody>
</table>

Figure 4. Percent SF over Time.

Student Data

Research Question 7: student academic achievement outcomes. TeachLivE simulations may contribute to changes in teacher practice, and effects might differ across teachers who received online PD as well, all of which may impact student academic outcomes. After removing missing cases resulting from students who did not complete both the pre- and posttest, a random sampling of 10% of the data were selected for analysis (n = 198). To investigate the impact on student outcomes, we performed a three-way mixed factors ANOVA with the factors of time (pre-intervention, post-intervention) × simulation (TeachLivE, no TeachLivE) × online PD (online PD, no online PD). The dependent variable was percentage correct on a 10-item student assessment based on the NAEP.
A three-way mixed ANOVA was conducted to understand the effects of TeachLivE, online PD, and time on percentage correct. Scores were not normally distributed, as assessed by Shapiro-Wilk’s test ($p < .05$); however, ANOVAs are considered to be robust to deviations from normality. There was homogeneity of variances for percentage correct at both pre- ($p = .645$) and post-intervention ($p = .598$), as assessed by Levene’s test for equality of variances. Results of the three-factor mixed design ANOVA indicated a differential effect of time for online PD when combined with TeachLivE ($F(1, 194) = 4.449, p = .036, \eta^2_p = .022$). Statistical significance of a simple two-way interaction was accepted at a Bonferroni-adjusted alpha level of .050. There was neither a statistically significant simple two-way interaction of online PD and TeachLivE at pre-intervention ($F(1, 194) = 2.370, p = .125, \eta^2_p = .012$), nor post-intervention ($F(1,194) = .087, p = .768, \eta^2_p = 0.00$). There was a main effect for time ($F(1, 194) = 14.043, p = .000, \eta^2_p = .068$). See Table 11 for changes in student scores pre- to post-test.

Table 11. Changes in Student Scores Pre- to Posttest

<table>
<thead>
<tr>
<th>Treatment Groups</th>
<th>n</th>
<th>M (SD) Pre</th>
<th>M (SD) Post</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison</td>
<td>50</td>
<td>53.78 (24.70)</td>
<td>53.00 (24.60)</td>
<td>-.78</td>
</tr>
<tr>
<td>Online PD</td>
<td>47</td>
<td>56.45 (26.73)</td>
<td>63.91(23.82)</td>
<td>+7.46</td>
</tr>
<tr>
<td>TeachLivE</td>
<td>52</td>
<td>48.85 (22.72)</td>
<td>59.02 (24.40)</td>
<td>+10.17</td>
</tr>
<tr>
<td>TeachLivE &amp; Online PD</td>
<td>49</td>
<td>62.24 (23.92)</td>
<td>67.96 (20.82)</td>
<td>+5.72</td>
</tr>
</tbody>
</table>

**Discussion**

In the present study, researchers investigated the use of the TeachLivE simulated classroom to increase HLPs, and whether taking online PD differentially increased those practices in both a simulated and real classroom. Further, changes in students’ achievement scores also were evaluated in real classrooms using questions from the NAEP for a pretest/posttest comparison.

Beginning with the classroom simulator, teachers overwhelmingly agreed that the classroom simulator felt like a real classroom and that the students also represented the kinds of students that existed in the real world. Further, our findings indicated that teachers asked significantly more DE questions and provided more SF to student-avatars in the simulator. That is, after four 10-minute sessions of TeachLivE, teachers increased their use of HLPs in the simulator, regardless of whether or not they had 40 minutes of additional online PD.

Results from the simulated classroom were reflected in the real classroom. In classes with real students, after four 10-minute sessions in TeachLivE, teachers asked significantly ($F(3, 130) = 3.479, p = .064, \eta^2_p = .026$) more DE questions than comparison groups, regardless of whether or not they had online PD. Although main effects for TeachLivE were not found for SF, TeachLivE combined with online PD produced a differential effect. Teachers who received TeachLivE without online PD had the highest percent of SF (+18%) across all four groups, while their counterparts who received online PD decreased their scores (-2%), the only decrease across all four groups. Despite a change, teachers’ SF observed in the classrooms should be interpreted...
with caution due to low reliability of observer scores. On a general measure of teacher performance in the classroom, TPOT sum, as with SF, teachers who received TeachLivE without online PD had the highest gains (+1.03) of the four treatment groups, yet their colleagues who received both TeachLivE and the online PD decreased (-.78), the largest decrease across all four groups. As predicted, teachers who received TeachLivE with no after-action-review on WT>3 did not differ significantly in their amount of WT. That is, by withholding feedback (after-action-review) from teachers after a simulation, their performance did not change.

Finally, in terms of student achievement data, all students’ scores increased significantly from pretest to posttest on 10 items from the NAEP assessment, which was expected as a result of instruction over the course of the year. However, differential effects of TeachLivE combined with online PD, seen in teachers’ SF and the general performance measure, were also echoed in student achievement scores. Students of teachers who received the combined treatment scored lower than those whose teachers received TeachLivE only.

As a whole, results from this study validate emerging research in the field that suggests that professional learning in mixed-reality simulated classrooms can be effective. We found support for our overarching hypothesis that virtual rehearsal in TeachLivE would increase teachers’ frequency of higher order questions and specific feedback to students, and that this increase also would be observed in their classrooms. Teachers who took part in a series of sessions in TeachLivE increased their instances of teaching practices in the simulator, similar to studies conducted earlier (e.g., Dawson & Lignugaris/Kraft, 2103; Elford et al., 2013; Vince Garland et al.; 2012). The current study extends the literature by demonstrating effects that extend HLPs for teachers from simulated classrooms to real classrooms.

The researchers hypothesized that online PD combined with TeachLivE would result in a differential increase, yet teachers who received both online PD and TeachLivE actually decreased their instances of HLPs. This finding was unanticipated. It is possible that although the content of the online PD and TeachLivE sessions were similar, teachers attempted to incorporate knowledge from the online PD into their virtual rehearsal, essentially working from a more complicated or divergent set of mental objectives for the simulation, resulting in a less effective simulation. Future research is needed to determine how to combine an array of PD that will make the strongest impact.

Limitations

The results should be considered in light of limitations to internal validity. Limitations resulted from the nested design in which teachers were grouped by school, because teachers within one school may be more similar than teachers across schools. Future research should include random assignment at the teacher level, rather than the school level, because performance in a simulator is individualized and threats to validity as a result of treatment diffusion (i.e., treatment effects spreading from one group to another) are unlikely. Random assignment at the teacher level would allow for balancing of similarities within each school. Further, the original research design was a group-randomized trial, but the nature of the design changed as a consequence of teacher requests to change initial treatment conditions, violating random assignment. In each case, teachers remained in the study but requested to downgrade
time commitment or reported that they would only participate if allowed to take part in the simulation. This activity changed the research design to a quasi-experiment.

Other threats to data reliability and confounding factors also existed. Classroom observation data had low reliability on SF and WT variables in the classroom. Data collectors were spread across the country, so data collection training should be improved in future studies to increase reliability of results. There may be confounding dispositions related to professional learning, computer simulation, or technology, especially related to online PD (which might explain interaction effects). Most significantly, classroom instruction was not standardized by a common lesson, and, as such, content and format varied widely. Future studies should take into account the need for a common lesson to provide a stronger basis for comparison.

As an intervention, delivery of TeachLivE requires moderate technology assets listed earlier. Also, the intervention is generally not delivered in the school setting, so teachers must travel to the simulation sites. Teachers receiving TeachLivE were required to visit the classroom simulator three times, which required significant scheduling efforts in the cases of last minute cancellations or delays resulting from technology issues. Future research should include plans for a mobile lab that could be brought to teachers’ classrooms, removing the barrier of teacher travel.

Future Research and Implications

Findings from this study can be generalized to other middle school mathematics teachers who receive four 10-minute sessions of TeachLivE with after-action-review. Teachers of other age levels and content areas should be considered in future research. Also, length and content of simulations should be varied to determine the optimal level of treatment needed to produce the desired results. Interaction of TeachLivE with other professional learning should be considered. Student achievement outcomes should be expanded to include a variety of measures to capture potential differences resulting from their teachers’ treatment; and most importantly, maintenance of effects over time should be considered.

The use of TeachLivE is being further investigated to determine if less time, additional sessions, or booster sessions would produce similar results or would maintain results over time. The ultimate goal of the research team is that the simulator does not replace “real” teaching but allows for safe and fun practice that is targeted and personalized. As new teachers enter the classroom, as teachers take leave and then come back to teaching, or when veteran teachers move into new roles, the hope is that simulators can be used to prepare and retool the skills of teachers at all levels from pre-service to in-service.

The team currently has three areas of unanswered questions related to time. First, if four 10-minute sessions impact practice, how long will this practice sustain? Do teachers need to practice once a month, a semester, a year, or every other year to ensure retention of new skills acquired in the simulator? Second, a pattern has been observed that after about five to seven minutes of working on a new skill, teachers tend to fall back to patterns of old behavior. Therefore, the team wonders if five to seven minutes of time in the simulator might be enough to impact practice, compressing the targeted PD activities to an even shorter period of time. What is the optimal session length needed to change a behavior? Third, how can the decoupling of
content and pedagogical teaching practices be best taught and taken apart and put back together? What amount of time is needed in each domain and how can the combination of these skills best impact the duality of skills all teachers need to make the strongest impact on student learning?

Beyond time, the research team also is considering if the simulator is best to work with teachers in an individual session, because we also have seen a great impact (but not a current part of our research) in using the simulator in a group setting. The widespread impact that could occur from a group session is yet to be measured, but the impact of individual versus group session format could further inform the field as more simulation technology is used in teacher education.

Just as results of group versus individual use need further investigation, so do methods of teacher feedback. Despite the integration of an after-action-review software in this study, the research team provided teachers with a handwritten summary of their data immediately after their TeachLivE session. The team plans to further investigate how feedback that is computerized may further enhance the simulated and personalized nature of TeachLivE.

With the agnostic nature of this simulator, we also want to consider the impact of this tool on other educational professionals such as administrators, guidance counselors, psychologists, and speech therapists. Beyond educational professional use in the simulator, the team at UCF has just started to investigate how this tool might also be used to impact student learning (student-to-avatar peer tutoring) and student life and social skill interactions with peer groups.

Simulated learning environments appear to provide an efficient tool for learning and practicing new teaching strategies, and four 10-minute simulator sessions on a specific teaching skill can change teacher behavior not only in the simulator with student-avatars, but also in the classroom with real students. Teachers have the opportunity to practice, make mistakes, and try new approaches to retool their teaching, all in a safe place for teachers and students. We believe TeachLivE is a disruptive technology and represents the next generation of professional learning and personalized learning for teachers. We plan to continue the work for teachers and with teachers, with the ultimate goal to directly impact student learning outcomes.
References


Appendix A

Teacher Practice Observation Tool
### Teacher Practice Observation Tool

**Establishing a culture for learning:**

- **Lack of commitment to learning**
- **Little commitment to learning**
- **High expectations by teacher**
- **Shared belief in importance**

**Engaging students in learning:**

- **Few engaged**
- **Some engaged**
- **Most engaged**
- **Virtually all highly engaged**

<table>
<thead>
<tr>
<th>Interval</th>
<th>Describe/ explain</th>
<th>Yes/No</th>
<th>3 or more</th>
<th>Less than 3</th>
<th>Specific</th>
<th>General</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wait Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Managing student behavior:**

- **No established standards of conduct**
- **Inconsistent standards of conduct**
- **Teacher established standards**
- **Students self-monitor with standards**

**Managing classroom procedures:**

- **Much instructional time is lost**
- **Some instructional time is lost**
- **Little loss of instructional time**
- **Time is maximized**

---

**Field Log**

**Establishing a culture for learning:**

- Write at least one note about practice per area then check appropriate box.

- **Lack of commitment to learning**
- **Little commitment to learning**
- **High expectations by teacher**
- **Shared belief in importance**

**Engaging students in learning:**

- **Few engaged**
- **Some engaged**
- **Most engaged**
- **Virtually all highly engaged**

---

**University: ___________**  
**Visit: 1 or 2 (circle one)**  
**School District _________**

**Teacher Code: ___________**  
**Date: _________**  
**Observer’s Initials: _________**

---

**Field Log 2 min:** Write at least one note about practice per area then check appropriate box.

---

**Sample**
## Teacher Practice Observation Tool

### Questions
1 min

### Wait Time
1 min

### Feedback
1 min

### Field Log
2 min: Write at least one note about practice per area then check appropriate box. Provide comments to justify rating.

### Communicating with students:
- explanations confusing or with errors
- explanations initially confusing
- explanations clearly communicated
- explanations clear & anticipate confusion

### Using questioning and discussion techniques:
- a few students respond
- some students discuss
- teacher engages most students
- students extend discussion

### Creating an environment of respect and rapport:
- mostly negative interactions
- generally appropriate interactions
- general caring and respect
- genuine warmth and caring

### Using assessment in instruction:
- little or none
- used sporadically
- used regularly
- fully integrated (formative assessment)

### Establishing a culture for learning:
- lack of commitment to learning
- little commitment to learning
- high expectations by teacher
- shared belief in importance

### Engaging students in learning:
- few engaged
- some engaged
- most engaged
- virtually all highly engaged
## Teacher Practice Observation Tool

### Questions
- **Interval:** 1 min
- **Description:** Describe/explain
- **Comments:**
  - Short
  - Yes/No
  - 3 or more
  - Less than 3
  - Specific
  - General

### Wait Time
- **Interval:** 1 min
- **Description:**
- **Comments:**
  - Short
  - Yes/No
  - 3 or more
  - Less than 3
  - Specific
  - General

### Feedback
- **Interval:** 1 min
- **Description:**
- **Comments:**
  - Short
  - Yes/No
  - 3 or more
  - Less than 3
  - Specific
  - General

### Field Log
- **Interval:** 2 min
- **Description:** Write at least one note about practice per area then check appropriate box. Provide comments to justify rating.

### Managing student behavior:
- **Description:**
  - no established standards of conduct
  - inconsistent standards of conduct
  - teacher established standards
  - students self-monitor with standards

### Managing classroom procedures:
- **Description:**
  - much instructional time is lost
  - some instructional time is lost
  - little loss of instructional time
  - time is maximized

### Communicating with students:
- **Description:**
  - explanations confusing or with errors
  - explanations initially confusing
  - explanations clearly communicated
  - explanations clear & anticipate confusion

### Using questioning and discussion techniques:
- **Description:**
  - a few students respond
  - some students discuss
  - teacher engages most students
  - students extend discussion

### Creating an environment of respect and rapport:
- **Description:**
  - mostly negative interactions
  - generally appropriate interactions
  - general caring and respect
  - genuine warmth and caring

### Using assessment in instruction:
- **Description:**
  - little or none
  - used sporadically
  - used regularly
  - fully integrated (formative assessment)
# Teacher Practice Observation Tool

<table>
<thead>
<tr>
<th>Interval</th>
<th>Describe/explain</th>
<th>Short</th>
<th>Yes/No</th>
<th>3 or more</th>
<th>Less than 3</th>
<th>Specific</th>
<th>General</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questions 1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wait Time 1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback 1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Field Log 2 min: Write at least one note about practice per area then check appropriate box. Provide comments to justify rating.

### Establishing a culture for learning:
- lack of commitment
- little commitment
- high expectations by teacher
- shared belief in importance

### Engaging students in learning:
- few engaged
- some engaged
- most engaged
- virtually all highly engaged

### Communicating with students:
- explanations confusing or with errors
- explanations initially confusing
- explanations clearly communicated
- explanations clear & anticipate confusion

### Using questioning and discussion techniques:
- a few students respond
- some students discuss
- teacher engages most students
- students extend discussion

### Creating an environment of respect and rapport:
- mostly negative interactions
- generally appropriate interactions
- general caring and respect
- genuine warmth and caring

### Using assessment in instruction:
- little or none
- used sporadically
- used regularly
- fully integrated (formative assessment)
Final Summary Check:
1. Now that you have completed the full observation, write brief statements of the teacher’s practice in the space provided below.
2. Choose the final level by checking **one** box within each category and be sure to provide justification for your rating.
3. If you are conducting this observation with another rater for inter-rater reliability, you may then discuss your chosen levels.
4. After your conversation, you may choose to make a change to one of the levels below.
5. If you choose to make a change, please provide a justification.
6. No changes may be made to the frequency counts or the field logs on pages 1-4.

<table>
<thead>
<tr>
<th>Managing student behavior:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ no established standards of conduct</td>
</tr>
<tr>
<td>□ inconsistent standards of conduct</td>
</tr>
<tr>
<td>□ teacher established standards</td>
</tr>
<tr>
<td>□ students self-monitor with standards</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Managing classroom procedures:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ much instructional time is lost</td>
</tr>
<tr>
<td>□ some instructional time is lost</td>
</tr>
<tr>
<td>□ little loss of instructional time</td>
</tr>
<tr>
<td>□ time is maximized</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Establishing a culture for learning:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ lack of commitment to learning</td>
</tr>
<tr>
<td>□ little commitment to learning</td>
</tr>
<tr>
<td>□ high expectations by teacher</td>
</tr>
<tr>
<td>□ shared belief in importance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Engaging students in learning:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ few engaged</td>
</tr>
<tr>
<td>□ some engaged</td>
</tr>
<tr>
<td>□ most engaged</td>
</tr>
<tr>
<td>□ virtually all highly engaged</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Communicating with students:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ explanations confusing or with errors</td>
</tr>
<tr>
<td>□ explanations initially confusing</td>
</tr>
<tr>
<td>□ explanations clearly communicated</td>
</tr>
<tr>
<td>□ explanations clear &amp; anticipate confusion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Using questioning and discussion techniques:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ a few students respond</td>
</tr>
<tr>
<td>□ some students discuss</td>
</tr>
<tr>
<td>□ teacher engages most students</td>
</tr>
<tr>
<td>□ students extend discussion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Creating an environment of respect and rapport:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ mostly negative interactions</td>
</tr>
<tr>
<td>□ generally appropriate interactions</td>
</tr>
<tr>
<td>□ general caring and respect</td>
</tr>
<tr>
<td>□ genuine warmth and caring</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Using assessment in instruction:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ little or none</td>
</tr>
<tr>
<td>□ used sporadically</td>
</tr>
<tr>
<td>□ used regularly</td>
</tr>
<tr>
<td>□ fully integrated (formative assessment)</td>
</tr>
</tbody>
</table>
Appendix B

Student Work Samples
Work Samples Packet

Thanks so much for your time today! We appreciate your time and effort. As you know we are looking for methods of professional development that are effective and efficient, and we thank you for your participation.

Enclosed in this packet are work samples from the avatar students you have just met. The work samples are based on an analysis of 113 actual student work samples in response to question one of the Assessment Task from the Lesson Plan for Solving Linear Equations in One Variable (page T-2).

You have just completed your first visit to the TeachLive™ classroom and met with our virtual students. You will have 4 more sessions of 10-minutes each in the simulator. Visit 1 was Session 1, Visit 2 will be Sessions 2 & 3, and Visit 3 will be Sessions 4 & 5.

For the remaining visits, you will practice teaching the Whole Class Introduction portion of the lesson (pages T-4 and T-5) and review how your practice changes over time. In each of the remaining sessions you will teach the content anew to the avatar students, an activity we call virtual rehearsal. The avatar students will not remember the content you have previously taught them, so you’ll get a chance to practice and improve your teaching skills.

For your next session, you’ll need to:

- Review all 5 student work samples enclosed
- Be prepared to facilitate as much of the Whole Class Introduction from the Lesson Plan for Solving Linear Equations in One Variable (for more information review pages T-1 through T-5) as you can. Don’t worry about covering all 15 minutes worth, since you only have 10 minutes in the simulator per session.
- Be familiar with the materials needed for the Whole Class Introduction (e.g., “True or False?” slides and “How many different values of x make the equation true?” slide), but do not bring these materials with you. The classroom will be equipped with the materials you need. When you enter the classroom, students will know they have completed the assessment task and will be ready for the Whole Class Introduction.

Again, thank you so much for your time!

You are paying it forward for teachers.
Too Good To Be True?

1. Amy and Ben are trying to answer this question:

When is the equation $5 - x = 6$ true?

They decide to compare their work.

Amy:

$5 - x = 6$

so $x = 6 - 5 = 1$

so it is true when $x = 1$

Ben:

You are taking a number away from 5

so the answer can’t be bigger.

Therefore it is never true.

Both Amy and Ben have made a mistake. Describe where they have gone wrong:

Amy: She did solve the equation. She should of’ add five to sit up like she did.

Ben: Ben did not do the opposite of substituting. He should of’ $+5$ to $5$ and to $c$

Show how you would work it out:

$5 - x = 6$

$4 + 15$

$x = 11$
Too Good To Be True?

1. Amy and Ben are trying to answer this question:

   When is the equation \( 5 - x = 6 \) true?

They decide to compare their work.

Amy:

\[
5 - x = 6 \\
so \quad x = 6 - 5 = 1 \\
so \text{ it is true when } x = 1
\]

Ben:

You are taking a number away from 5 so the answer can't be bigger. Therefore it is never true.

Both Amy and Ben have made a mistake. Describe where they have gone wrong:

Amy: Amy is wrong by saying \( x = 1 \) because if \( x = 1 \) it would be \( 5 - 1 = 6 \) and that makes it impossible for the outcome to be 6.

Ben: Ben is correct by saying the outcome could never be true.

Show how you would work it out:

You take 11 in place of \( x \) and substitute it by 5. You would get 6 as the outcome. \( 5 - 11 = 6 \) true.
1. Amy and Ben are trying to answer this question:

When is the equation $5 - x = 6$ true?

They decide to compare their work.

Amy:

$$5 - x = 6$$

so $x = 6 - 5 = 1$

so it is true when $x = 1$

Ben:

You are taking a number away from 5
so the answer can't be bigger
Therefore it is never true.

Both Amy and Ben have made a mistake. Describe where they have gone wrong:

Amy:

She made a mistake because
the number $x$ is equal to Should be negative.

Ben:

It is possible to get a larger number
when taking a number away from 5.

Show how you would work it out:

$$5 - x = 6$$

5 - S - S - x = 1

$x = 1$
Too Good To Be True?

1. Amy and Ben are trying to answer this question:

When is the equation \( 5 - x = 6 \) true?

They decide to compare their work.

Amy:

\[
5 - x = 6 \\
so \ x = 6 - 5 = 1 \\
so \ it \ is \ true \ when \ x = 1
\]

Ben:

You are taking a number away from 5
so the answer can't be bigger
Therefore it is never true.

Both Amy and Ben have made a mistake. Describe where they have gone wrong:

Amy: If you subtract 1 from 5, you don't get 6.

Ben: The answer can be bigger, you will just get a negative.

Show how you would work it out:

\[
\begin{align*}
\text{Amy:} & \quad 5 - x = 6 \\
& \quad 5 - 6 = x \\
& \quad x = -1
\end{align*}
\]

\[
\begin{align*}
\text{Ben:} & \quad 5 - x = 6 \\
& \quad 5 - 6 = x \\
& \quad x = -1
\end{align*}
\]
1. Amy and Ben are trying to answer this question:

When is the equation $5 - x = 6$ true?

They decide to compare their work.

**Amy**

$5 - x = 6$
so $x = 6 - 5 = 1$
so it is true when $x = 1$

**Ben**

You are taking a number away from 5
so the answer can’t be bigger.
Therefore it is never true.

Both Amy and Ben have made a mistake. Describe where they have gone wrong:

Amy: She should have done $5 - x = 6$ to figure out the answer. Amy did the problem wrong. She shouldn’t have done $6 - 5 = 1$. You can’t change the problem.

Ben: Ben is wrong.

Show how you would work it out:

____________________________________________________________________

____________________________________________________________________

____________________________________________________________________