# Using Simulated Virtual Environments to Improve Teacher Performance

Lisa A. Dieker
University of Central Florida

Charles E. Hughes
University of Central Florida

Michael C. Hynes
University of Central Florida

Carrie Straub
University of Central Florida

Abstract: National research that is a collaborative between universities and school districts is critical to ensure innovative ideas are created to directly impact teacher performance in the classroom. This paper describes a national research study on using simulation in teacher professional development to impact teacher practice. A quasi-experimental, pre-post group design was used to examine the effects of the simulator on middle school teachers' practices in mathematics. 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. Results from this study validate emerging research in the field of teacher professional development and simulation that suggests that professional learning in mixed-reality simulated classrooms can be effective in impacting teacher practice.

**KEYWORDS:** Danielson Framework for Teaching, preservice teachers, Professional Development Schools, High Leverage Practices, simulations, TeachLivE, teacher performance, teacher preparation, virtual environments

**NAPDS NINE ESSENTIALS ADDRESSED:**
2. A school–university culture committed to the preparation of future educators that embraces their active engagement in the school community;
3. Ongoing and reciprocal professional development for all participants guided by need; and
4. A shared commitment to innovative and reflective practice by all participants

Teachers are the single most important factor to influence student learning and academic outcomes, aside from the students themselves (Darling-Hammond, 2003; Kane & Staiger, 2008). High quality professional development (PD) is crucial for teachers to meet the new levels of standards in today's classrooms. The ultimate outcome of any PD is to make a positive impact on teacher practice and student academic outcomes. Due to the complex nature of collecting student data in schools (Guskey & Sparks, 2002), research on PD using the What Works Clearinghouse (WWC) standards on student achievement is limited (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 nine 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 identified studies cited active learning and opportunities for teachers to adapt practices to their individual classrooms as having the greatest impact. Lauer, Christopher, Firpo-Triplett, and Buchting (2014) in their review of the literature found that 30 hours was a key point for PD to be effective but that it had to be grounded in teacher practice.

The search for effective practices and what should be included in effective PD to impact student learning (Earley & Porritt, 2014) emerged from a national study on the Measures of Effective Teachers (http://www.metproject.org/). This group reported as Teaching Works (2014; http://www.teachingworks.org/) analyzed core practices for teachers and developed a set of 19 high-leverage practices (HLPs) for instruction across content areas, that most likely lead to increased advances in student learning. These 19 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). These practices were created from a "thorough analysis of what teachers do and yielded a large 'map' of nearly 100 tasks of teaching" (Teaching Works, 2014, para. 1) and then narrowed to 19 practices by expert teachers. The result was the creation of a framework of synthesized teaching practices aligned with the highest student learning outcomes. These 19 practices can be found at http://www.teachingworks.org/work-of-teaching/high-leverage-practices and are intended to provide a framework for skills that should be targeted in high-quality teacher education and professional development no matter the context (e.g., face-to-face, online, simulation) in which the professional development is being delivered.

Similar teaching practices are described in other published descriptions of teacher practice such as the Danielson framework. This framework is a "research-based set of component of instruction, aligned to the INTASC standards, and grounded in a constructive view of learning and teaching" (Danielson, 2011, p. v) and consists of four domains related to teacher practice to impact student learning. These domains provide a framework for teachers to discuss their practices and to identify areas in need of further enhancement in their own instruction. The Danielson framework provides indicators for eliciting student thinking, such 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").

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 this PD to teachers? These frameworks, such as the Danielson and the HLPs, should be paired with practices such as those outlined by the National Association of Professional Development Schools (NAPDS) to provide the school and university partners with new and innovative models of PD. Yet, these models must be grounded in teacher practices that are aligned with increasing student learning outcomes.

Examples of next generation PD environments for teachers to learn both pedagogical and content skills aligned with recognized national frameworks are emerging, and computer simulation is at the forefront. As described in Dieker, Straub, Hughes, Hynes, and Hardin, (2014), simulated environments can provide educational experiences and opportunities that may not be available in real-world settings (Dieker, Straub et al., 2014; Dieker, Rodriguez, Lignugaris/Kraft, Hynes, & Hughes, 2014) and allow for safe practice of targeted skills (no "real" students are put at risk) until mastery is achieved. Despite this emergence of a new tool, the ultimate question that has to be answered is, "Does teacher professional development in virtual environments transfer to practice and impact student learning?"

## A Virtual Learning Environment

This study is the first national study conducted using a virtual learning environment in teacher education and PD. The specific environment used here, TLE TeachLivE™ (TeachLivE), is an immersive, virtual classroom simulator that includes the features of a real classroom (see http://www.teachlive.org). Real and virtual worlds are combined in the environment to give the users a sense of immersion with both physical and social presence (Biocca, Harms, & Burgoon, 2003; Hayes, 2015), wherein 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 content knowledge, discourse, and behavioral compliance. As more and more universities are using this virtual simulation tool, a research base is emerging focusing on the use of TeachLivE with teacher candidates and teachers.

A unique component of the TeachLivE environment is the opportunity to offer standardized experiences grounded in the HLPs with integrated video tagging software to record, play back, and export data collected during a session. That is, the simulated activity allows time for an integrated after-action-review (AAR) process to take place, in which teachers take part in structured reflection (Baird, Holland, & Deacon, 1999). Time in a simulator is compressed so that 10 minutes equates to between 30 and 60 minutes of real time (we can skip the starting, getting students seated or, if desired, start in the middle of a lesson with student work samples already produced). Most importantly, unlike in real classrooms, teachers can re-enter the simulated environment to fix instructional "errors" 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 and effectiveness of the use of simulation for teacher education.

## Theoretical Framework and Overarching Hypotheses

The overarching hypothesis of this study is that teachers who engage in virtual PD will improve their application of pedagogical knowledge as well as improve student content knowledge. The research team hypothesized that teacher learning is most effective in contextually meaningful settings (Dieker, Rodriguez, et al., 2014) and created a contextually meaningful simulation activity that provided learners with the opportunity to practice HLPs with student-avatars. This work is grounded in Brown, Collins, and Duguid's (1989) theory of situated cognition, asserting "what is learned cannot be separated from how it is learned and used" (p. 88). Further, the research team hypothesized that learning occurring in a virtual classroom would transfer to a real classroom.

## Research Questions for Teacher Performance

As noted in the hypothesis, the focus of this research study was on changing teacher practice. The team set about finding evidence of change in teacher practice in two environments: (a) the classroom simulator and (b) the teachers' classrooms. In both settings the team attempted to change teacher practice using TeachLivE or TeachLivE combined with other forms of PD. Research questions focused on the effect of the classroom simulator environment as compared to lesson resources and synchronous online PD on teacher performance (both in the virtual environment and back in their real classroom) and student learning outcomes. The researchers grounded the study in the following research questions: (1) 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?; (2) What are the effects of simulation without after-action-review on teaching practice in a classroom?; (3) Are there differential effects of TeachLivE on teacher practice in a classroom based on whether or not teachers received online PD?; and (4) Are there differential effects of TeachLivE on student scores based on whether or not their teachers received online PD? These research questions were explored by working with middle school teachers in 10 research locations comprised of university and school district partners.

## Method

### Subjects

Participants were practicing middle school mathematics teachers. At each of 10 partnership sites, approximately 20 teachers were self-nominated or nominated by their supervisors with the intent of receiving innovative, technology-rich PD. Of those teachers, 135 teachers completed the study. Participation was entirely voluntary with minimal to no compensation provided. Demographic data for participating teachers are presented in Table 1.

Data were collected in two settings: the teachers' real classrooms (before and after the simulator experiences) and the classroom simulator. Teachers were observed in their respective middle school classrooms located in six states: Florida, Louisiana, Michigan, Mississippi, New York, and Utah. School settings included urban, suburban, and rural with public or private enrollment. Virtual simulated classrooms were located at 10 partner sites across the country in rooms at a nearby university or school district partner sites.

**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, a validated tool used to measure teacher practice; Hayes, Hardin, Dieker, Hynes, Hughes, & Straub, 2013).

Table 1. Teacher Demographic Data

| | Control (*n* = 35) | | PD Only (*n* = 35) | | TLE Only (*n* = 35 | | TLE & PD (*n* =30 | |
|---|---|---|---|---|---|---|---|---|
| Variable | *n* | (%) | *n* | (%) | *n* | (%) | *n* | (%) |
| Professional licensure | | | | | | | | |
| Yes | 32 | (91) | 32 | (91) | 31 | (89) | 29 | (97) |
| No | 0 | (0) | 1 | (3) | 0 | (0) | 1 | (3) |
| No response | 3 | (9) | 2 | (6) | 4 | (11) | 0 | (0) |
| If licensed, is license in math? | | | | | | | | |
| Yes | 26 | (74) | 25 | (71) | 25 | (71) | 26 | (87) |
| No | 6 | (17) | 7 | (20) | 6 | (17) | 4 | (13) |
| No response | 3 | (9) | 3 | (9) | 4 | (11) | 0 | (0) |
| Area of certification | | | | | | | | |
| Grades 5-9 math only | 13 | (37) | 15 | (43) | 9 | (26) | 10 | (33) |
| Grades 6-12 math only | 6 | (17) | 8 | (23) | 12 | (34) | 11 | (37) |
| Other | 9 | (26) | 3 | (9) | 6 | (17) | 4 | (13) |
| Grades 5-9 & 6-12 math | 1 | (3) | 0 | (0) | 1 | (3) | 2 | (7) |
| Grades 5-9 math & other | 0 | (0) | 1 | (3) | 1 | (3) | 1 | (3) |
| Grades 6-12 math & other | 0 | (0) | 1 | (3) | 0 | (0) | 1 | (3) |
| Grades 5-9 & 6-12 math, & other | 1 | (3) | 0 | (0) | 0 | (0) | 0 | (0) |
| No response | 5 | (14) | 7 | (20) | 6 | (17) | 1 | (3) |
| Grade levels taught | | | | | | | | |
| 6-8 only | 17 | (49) | 14 | (40) | 18 | (51) | 19 | (63) |
| K-5 & 6-8 | 5 | (14) | 7 | (20) | 6 | (17) | 3 | (10) |
| 6-8 & 9-12 | 6 | (17) | 10 | (29) | 7 | (20) | 8 | (27) |
| K-5, 6-8, & 9-12 | 4 | (11) | 2 | (6) | 0 | (0) | 0 | (0) |
| No response | 3 | (9) | 2 | (6) | 4 | (11) | 0 | (0) |
| Highest academic level | | | | | | | | |
| Bachelor's | 17 | (49) | 19 | (54) | 18 | (51) | 21 | (70) |
| Master's | 15 | (43) | 14 | (40) | 13 | (37) | 9 | (30) |
| No response | 3 | (9) | 2 | (6) | 4 | (11) | 0 | (0) |
| Area of masters degree | | | | | | | | |
| Math education | 2 | (6) | 4 | (11) | 4 | (11) | 2 | (7) |
| Educational leadership | 5 | (14) | 2 | (6) | 2 | (6) | 2 | (7) |
| Other | 7 | (20) | 9 | (26) | 7 | (20) | 7 | (23) |
| Educational leadership & other | 1 | (3) | 0 | (0) | 0 | (0) | 0 | (0) |
| Not applicable | 14 | (40) | 13 | (37) | 14 | (40) | 16 | (53) |
| No response | 6 | (17) | 7 | (20) | 8 | (23) | 3 | (10) |
| Years teaching math | | | | | | | | |
| One year | 2 | (6) | 6 | (17) | 7 | (20) | 3 | (10) |
| Two years | 2 | (6) | 4 | (11) | 2 | (6) | 5 | (17) |
| Three years | 2 | (6) | 1 | (3) | 5 | (14) | 1 | (3) |
| Four years | 2 | (6) | 2 | (6) | 0 | (0) | 4 | (13) |
| 5-10 years | 13 | (37) | 9 | (26) | 9 | (26) | 9 | (30) |
| More than 10 years | 11 | (31) | 11 | (31) | 8 | (23) | 8 | (27) |
| No response | 3 | (9) | 2 | (6) | 4 | (11) | 0 | (0) |

| | Control (*n* = 35) | | PD Only (*n* = 35) | | TLE Only (*n* = 35 | | TLE & PD (*n* =30 | |
|---|---|---|---|---|---|---|---|---|
| Variable (Cont's) | *n* | (%) | *n* | (%) | *n* | (%) | *n* | (%) |
| Age | | | | | | | | |
| 18-29 | 7 | (20) | 1 | (3) | 7 | (20) | 7 | (23) |
| 30-39 | 11 | (31) | 7 | (20) | 10 | (29) | 9 | (30) |
| 40-49 | 6 | (17) | 12 | (34) | 6 | (17) | 11 | (37) |
| 50 or above | 8 | (23) | 13 | (37) | 8 | (23) | 3 | (10) |
| No response | 3 | (9) | 2 | (6) | 4 | (11) | 0 | (0) |
| Gender | | | | | | | | |
| Male | 7 | (20) | 9 | (26) | 8 | (23) | 5 | (17) |
| Female | 25 | (71) | 24 | (69) | 23 | (66) | 25 | (83) |
| No response | 3 | (9) | 2 | (6) | 4 | (11) | 0 | (0) |
| Ethnicity | | | | | | | | |
| American Indian | 0 | (0) | 0 | (0) | 0 | (0) | 1 | (3) |
| Asian | 2 | (6) | 1 | (3) | 0 | (0) | 1 | (3) |
| Black | 2 | (6) | 3 | (9) | 3 | (9) | 1 | (3) |
| Hispanic | 4 | (11) | 2 | (6) | 0 | (0) | 0 | (0) |
| White | 24 | (69) | 26 | (74) | 28 | (80) | 27 | (90) |
| Other | 0 | (0) | 1 | (3) | 0 | (0) | 0 | (0) |
| No response | 3 | (9) | 2 | (6) | 4 | (11) | 0 | (0) |

Valid Participants = 135 Attrition = 22

During classroom observations, data were collected on (a) the frequency of HLPs determined to increase the likelihood that these teaching behaviors would have a positive effective on students' learning outcomes (Teaching Works, 2014), (b) modified sub-constructs from the 2011 Danielson Framework for Teaching Evaluation Instrument, and (c) qualitative field notes. The HLPs provided the defined behaviors to be observed and were aligned with the study at the request of the funding agency. The Danielson framework provided sub-constructs for observation of practices that overlapped with the HLPs and overall teacher development. Data were collected in five-minute intervals, rotating across constructs, so observers focused on one construct at a time during the interval. For the teachers who experienced the classroom simulator, data also were collected on their four sessions in the TeachLivE environment, and the frequency of the specific targeted HLPs exhibited in each session in the simulated classroom were coded.

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 (a) asking student to describe or explain types of questions; (b) short response questions; and (c) yes/no questions. These targeted HLP's were selected based upon feedback from teacher preparation experts involved with the study and were identified as both observable and measurable behaviors that could be seen in the simulator. The final reason for selection was that the mathematic experts involved with the study felt these behaviors aligned with best practices in math discussion and also aligned with both the HLP's and the Danielson framework.

In the classroom simulator, frequency data were collected. Each session lasted 10 minutes and focused on teachers' discussing students' error patterns in mathematics from already completed student work samples. In the simulator the frequency and type of instances for each behavior were noted. In the teachers' classrooms, lessons varied in length (45 to 95 minutes), so a percentage of describe/explain questions was calculated based on the ratio of occurrences of describe/explain questions to the sum of all questions (describe/explain, short, and yes/no) asked by the teacher.

Frequency data were coded on 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 or general feedback.

Just as with other teacher behaviors 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 based on the ratio of the occurrences of specific feedback to all feedback (specific plus general).

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.

**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 (MET 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 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. The research team in an earlier study (Straub, Dieker, Hynes, & Hughes, 2014) created a rating scale of what was observable for each of these four constructs with level 1 being limited observation of the skill to level 4 being mastery of the skill. These categories were vetted, validated, and reliability training occurred for the research team prior to use in classroom observations and simulation observations.

**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. 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 group 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 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.
During each session in the simulator, videos were tagged for frequency and type of questions and feedback. See Table 2. Dependent Variables for an overview of dependent variables collected in the study.

**Student data.** Data also were collected from middle school students in the participating teachers' classrooms on student performance on a curriculum-based measure of algebraic equations based on the National Assessment of Educational Progress (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 (pre-post).
Table 2. Dependent Variables

| Construct | Type |
|---|---|
| **HLP # Questioning** | Describe/explain questions<br>Short response questions<br>Yes/no questions<br>Wait time greater than or equal to three seconds<br>Wait time less than three seconds |
| **HLP #12 Feedback** | Specific feedback<br>General feedback |
| **Sub-constructs from Danielson Framework** | Establishing a culture for learning<br>Engaging students in learning<br>Managing student behavior<br>Managing classroom procedures<br>Communicating with students<br>Using questioning and discussion techniques<br>Creating an environment of respect and rapport<br>Using assessment in instruction |

## 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 (G1 36, G2 41, G3 41, G4 39) were grouped together into a unit and randomly assigned to one of four treatment conditions as a unit. The random assignment procedure took place at all 10 partnership sites, resulting in four experimental groups.

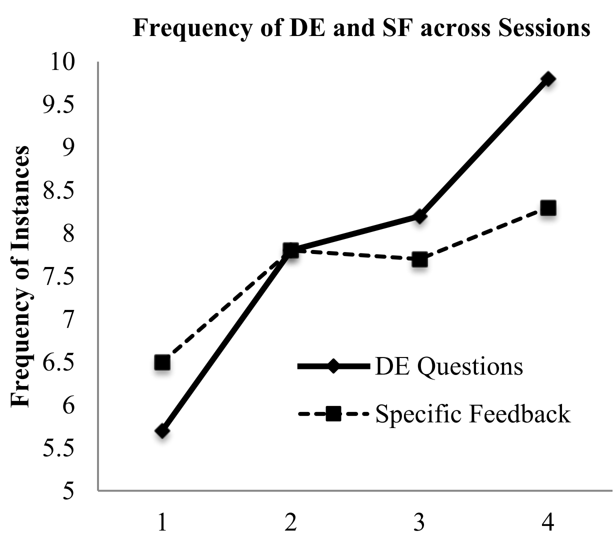
## Interventions

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 (G1) teachers served as a comparison group and received lesson plans only; Group 2 (G2) teachers received lessons and online PD; Group 3 (G3) teachers received lessons and TeachLivE; and

Group 4 (G4) teachers received lessons, online PD, and TeachLivE (without AAR). See Figure 1 for an overview of the four treatment groups. As with all 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 school year. They were given no other intervention while in this study.

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 expert in the Classroom Challenges curriculum, delivered via the Adobe Connect platform. 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 (the same samples used in the simulator) in response to a formative assessment included in the lesson 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 to be spent in the simulator.

Figure 1. Frequency of Describe Explain and SF across sessions




Teachers in G3 received a digital copy of the lesson (like teachers in G1 and G2), as well as four 10-minute virtual sessions in TeachLivE. In the simulator, teachers attended individual PD and interfaced with TeachLivE. Classroom simulators at the 10 partner sites across the country were connected via secure server and provided fidelity of treatment as all sessions were controlled from the primary research site. For operation at the teacher partner 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. 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 orientation sessions, but here those sessions were in the TeachLivE system. After orientation, teachers received four 10-minute sessions in TeachLivE as PD with data on targeted behaviors gathered during each session. Prior to any of the sessions, teachers were given the same student work samples used in the online PD, but in this condition, teachers were told that each work sample was a product of a specific student-avatar. Teachers were instructed to teach 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.

Teachers in G4, the TeachLivE and Online PD group, received the lesson plan, participated in the online PD, and engaged in virtual teaching in TeachLivE as outlined above. However, they did not receive any after-action review.

The 135 teachers were grouped by school, then randomly assigned to four groups in a randomized group design nested within school. 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.

## 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. The online PD was monitored by a facilitator who 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 Analyses

Teaching practices were defined on three distinct dimensions of pre- and post-intervention: (a) describe/explain questions (DE), (b) specific feedback (SF), and (c) summary score on the TPOT (TPOT Sum). For research question 1, to determine if a difference in effects in performance occurred from four 10-minute sessions of TeachLivE in a classroom simulator based on whether or not teachers received 40 minutes of online PD, 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 and SF. Due to a lack of research in simulation (as similar research in other fields has not occurred related to performance due to the life and death nature of past use of simulation – pilot training, surgery, military training) the team predetermined that a level of significance with greater type I error would be considered, as the simulator creates an environment without risk and any changes of teacher performance provided a foundation for future research. Partial eta squared was used to interpret effect size rather than eta squared because a multifactor design was used (Pierce, Block, & Aguinis, 2004), and there was a desire by the team to compare effects across different factorial designs used in the study (Levine & Hullet, 2002).

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 ($F(3,171) = .735$, $p = .532$, $\eta 2p = .13$). However, a significant time effect was identified ($F(3,171) = 9.993$, $p = .000$, $\eta 2p = .149$). Pallant (2007) recommends interpreting partial eta squared using Cohen's (1988) guidelines for eta squared effect size: small (.01), medium (.06), or large (.14), resulting in a finding of a large effect for time (i.e., session). Mean scores increased at each session (see Table 3).

Table 3. Mean DE Questions across 10-minute TeachLivE Sessions

| | | TeachLivE Sessions | | | |
|---|---|---|---|---|---|
| | | Session 1 | Session2 | Session 3 | Session 4 |
| PD Factor | *n* | M *(SD)* | M *(SD)* | M *(SD)* | M *(SD)* |
| No Online PD | 34 | 5.1 *(4.2)* | 7.6 *(4.9)* | 8.4 *(5.3)* | 9.9 *(4.8)* |
| Online PD | 25 | 6.5 *(4.2)* | 7.9 *(4.5)* | 7.9 *(5.9)* | 9.5 *(8.1)* |
| Total | 59 | 5.7 *(4.2)* | 7.8 *(4.7)* | 8.2 *(5.5)* | 9.8 *(6.3)* |

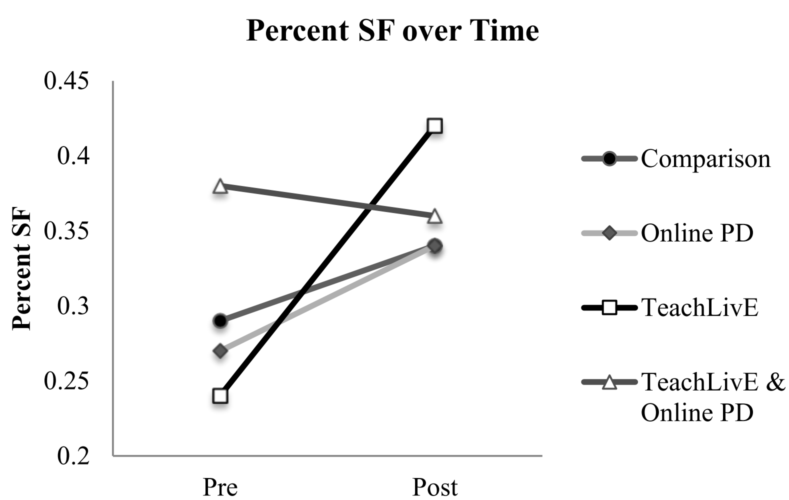
Two observers collected data on frequency of SF given to student-avatars by teachers per TeachLivE session. Reliability of scores between observers during each session was calculated (session 1, r = .928; session 2, r = .872; session 3, r = .811; session 4, r = .790). Results from a two-factor mixed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

Table 4. Mean SF Questions across 10-minute TeachLivE Sessions

| | | TeachLivE Sessions | | | |
|---|---|---|---|---|---|
| | | Session 1 | Session2 | Session 3 | Session 4 |
| PD Factor | *n* | M *(SD)* | M *(SD)* | M *(SD)* | M *(SD)* |
| No Online PD | 34 | 6.2 *(5.1)* | 8.3 *(6.0)* | 8.7 *(4.8)* | 8.6 *(4.6)* |
| Online PD | 24 | 6.9 *(4.5)* | 6.7 *(3.6)* | 6.3 *(3.9)* | 7.9 *(6.7)* |
| Total | 58 | 6.5 *(4.8)* | 7.8 *(5.1)* | 7.7 *(4.6)* | 8.3 *(5.5)* |

To investigate the effects on teacher practice in a classroom setting, observers collected data during classroom observations pre- and post-treatment. Teacher behavior was considered without and with an integrated after-action-review process in TeachLivE. To examine performance of teachers in a classroom after TeachLivE sessions without after-action-review, a three-factor mixed design ANOVA was calculated 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).

Figure 2. Mean Scores of Frequency of Instances across Sessions




*Note G4 TeachLivE & Online PD did not receive AAR*

The dependent variable was percentage of wait time that was three seconds or more (WT>3). 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. 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 non-significant effect for the three-way interaction effects of time, simulation, and online PD ($F(1,130) = 1.003$, $p = .318$, $\eta 2p = .008$). No effects were found for simple two-way interaction between time and simulation ($F(1,130) = .002$, $p = .968$, $\eta 2p = .000$), and this finding was expected because no performance feedback had been provided to teachers. Further, no effects were found for simple two-way interaction between time and online PD ($F(1,130) = .304$, $p = .582$, $\eta 2p = .002$) or for time ($F(1,130) = 1.580$, $p = .211$, $\eta 2p = .012$).

While no significant effects for TeachLivE simulation without after-action-review were found, TeachLivE with after-action-review did contribute to changes in teacher practice, and that effect differed across teachers who received online PD as well. Again, the research team 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.

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 ($p < .05$). 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 ($F(1,130) = .168$, $p = .682$, $\eta 2p = .001$). The interaction between TeachLivE and online was not statistically significant ($F(1,130) = .015$, $p = .902$, $\eta 2p = .000$). There was a statistically significant two-way interaction between time and online PD ($F(1,130) = 5.735$, $p = .018$, $\eta 2p = .042$) and time and TeachLivE ($F(1,130) = 3.479$, $p = .064$, $\eta 2p = .026$).

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 that was considered a family of comparisons. Adjusted p-values are reported. Statistically significant differences existed at pre-intervention for those assigned to online PD, ($F(1,130) = 4.854$, $p = .029$, $\eta 2p = .036$), but not at post-intervention ($F(1,130) = 1.204$, $p = .902$, $\eta 2p = .275$), 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), $p = .029$. 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 ($F(1,130) = 1.274$, $p = .261$, $\eta 2p = .010$), but there was a post-intervention ($F(1,130) = 9.827$, $p = .002$, $\eta 2p = .070$), suggesting 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 receive online PD 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 5 for mean changes from pre to post.

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 ($M = 14\%$).

Table 5. Means Changes in Percent DE.

| | TeachLivE Factor | | Online PD Factor | |
|---|---|---|---|---|
| | TeachLivE | No TeachLivE | Online PD | No Online PD |
| Time Factor | M *(SD)* | M *(SD)* | M *(SD)* | M *(SD)* |
| Pre | 18 *(17)* | 16 *(14)* | 20 *(16)* | 14 *(13)* |
| Post | 24 *(20)* | 14 *(16)* | 17 *(18)* | 21 *(20)* |
| Change | +6 | -2 | -3 | +7 |

Next, SF was evaluated. Pearson's correlation provided a basis for interpreting reliability of scores between observers. 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 ($p < .05$). 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 ($F(1,130) = 3.486$, $p = .064$, $\eta 2p = 0.26$). 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 ($F(1, 131) = 3.638$, $p = .059$, $\eta 2p = .027$), but not at post-intervention ($F(1,130) = .527$, $p = .469$, $\eta 2p = .004$). 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. No significant positive changes were found between G1 and G2 so no further analysis of these groups were completed. 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.

Finally, TPOT sum scores were evaluated. 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 ($p < .05$). 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 ($F(1,117) = 3.003$, $p = .086$, $\eta 2p = .025$). 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, $\eta 2p$ = .009), nor post-intervention ($F$(1,121) = .008, $p$ = .928, $\eta 2p$ = .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 6 for changes in scores over time.

Table 6. Changes in Mean Score of TPOT Sum over Time.

| | | Time | | |
|---|---|---|---|---|
| | | Pre | Post | |
| Treatment Groups | *n* | M *(SD)* | M *(SD)* | Change |
| Comparison | 32 | 22.06 *(3.75)* | 22.00 *(4.20)* | - .06 |
| Online PD | 32 | 21.33 *(5.35)* | 21.83*(4.81)* | + .50 |
| TeachLivE | 32 | 21.63 *(4.53)* | 22.66 *(3.97)* | +1.03 |
| TeachLivE & Online PD | 27 | 23.19 *(3.88)* | 22.41 *(4.49)* | - .78 |

## Discussion

In the present study, researchers investigated the use of the TeachLivE simulated classroom to increase HLPs (Loewenberg Ball, & Forzani, 2010), 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. The use of this PD model aligned with teacher performance being embedded in a true relationship between a university and local school districts. This type of collaborative partnership aligns with the recommended practices associated with the NAPDS. This study specifically aligns with four of the nine practices found in effective school-based PD of "2) A school–university culture committed to the preparation of future educators that embraces their active engagement in the school community; 4) A shared commitment to innovative and reflective practice by all participants; 7) A structure that allows all participants a forum for ongoing governance, reflection, and collaboration; and 8) Work by college/university faculty and P–12 faculty in formal roles across institutional settings" (NAPDS, 2008). A recommendation to the field would be to further expand work with online PD and simulation to align with all 8 components in creating long-term partnerships for teacher PD embedded in PDS.

This future embedding of simulation aligns with the findings of this research team. In this study, teachers overwhelmingly agreed that the classroom simulator felt like a real classroom and that the avatar students represented the kinds of students that existed in the real world. Further, teachers asked significantly more DE questions and provided more SF to avatars as sessions progressed. 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. This immediate and transferable impact of skills makes the need for ongoing PDS partnerships to align with these types of "innovative" practices essential.

Results from the simulated classroom were reflected in the real classroom after four 10-minute sessions in TeachLivE. In classes with real students, teachers asked significantly ($F(3,130) = 3.479$, $p = .064$, $\eta 2p = .026$) 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 across all four groups, while their counterparts who received online PD decreased their scores, the largest decrease across all four groups. Although this change could not be explained, this finding does provide an opportunity for further research. On a general measure of teacher performance in the classroom, all teachers improved significantly from pre- to post-observation. As with SF, teachers who received TeachLivE without online PD had the highest gains (+1.03) of the four treatment groups. As predicted, teachers who received TeachLivE with no after-action-review on WT>3 did not show significant improvment 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, also were echoed in the student achievement scores.

Taken as a whole, results from this study support emerging research in the field that suggests that professional learning in virtual environments can be an effective tool for PD that transfers to classroom practice. The researchers found support for the overarching hypothesis that time in the virtual environment increases teachers' frequency of higher order questions and specific feedback to students, and that this increase also was 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 smaller studies conducted earlier (e.g., Dawson & Lignugaris/Kraft, 2103; Elford, James, & Haynes-Smith, 2013; Vince Garland et al., 2012). The current study contributes to the literature by demonstrating effects that extend HLPs for teachers from simulated classrooms to real classrooms.

## 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. Random assignment at the teacher level would allow for balancing of similarities within each school.

As an intervention, delivery of TeachLivE requires moderate technology assets (computer, projection screen, projector and a Kinect). 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 the use of a mobile lab 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 and aligned specifically with recommended practices for best practices in university-school partnerships (NAPDS, 2008). 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 (Earley & Porritt, 2014).

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, improving on the practice of requiring 30 hours to potentially produce teacher learning gains (Lauer et al., 2014). The ultimate goal of the research team is not to replace "real" teaching with the simulator, but rather to use the simulator to allow for safe practice that is targeted and personalized. The team also hopes to build upon the concern by Guskey and Yoon (2009), to find effective learning opportunities through virtual environments to provide PD that teachers can immediately adapt for their individual classrooms to positively impact student learning. 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 for future research related to time. First, if four 10-minute sessions impact practice, how long will this practice sustain? Are periodic sessions required to ensure retention of new skills acquired in the simulator? Second, what is the optimal session length needed to change a behavior? Third, how can the decoupling of content and pedagogical teaching practices best be taught and taken apart and put back together? The research team plans to continue the work for teachers and with teachers, with the ultimate goal to directly impact student learning outcomes.

With the agnostic nature of this simulator, the research team also wants to consider the impact of this tool on other educational professionals such as administrators, guidance counselors, psychologists, and speech therapists. The future of research in teacher education has the potential for more standardized approaches and comparison through work with simulation, much like is seen in flight simulation training or standardized patient care in medicine, providing safe ways to help shape and support teachers in targeted areas of concern.

## References

Adobe Systems Incorporated. (2014). Adobe Connect. [Computer software]. Santa Clara, CA: Adobe Systems Incorporated.

Baird, L., Holland, P., & Deacon, S. (1999). Learning from action: Imbedding more learning into the performance fast enough to make a difference. Organizational Dynamics, 27(4), 19-32.

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. Presence: Teleoperators and Virtual Environments, 12(5), 456-480. doi: 10.1162/105474603322761270

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 328-375). New York, NY: Macmillan.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18(1), 32-42. doi:10.3102/0013189X018001032

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Common Core Standards Initiative. (2011). Preparing America's students for college and career. Retrieved from http://www.corestandards.org

Danielson, C. (2011). The Framework for Teaching Evaluation Instrument. In T. D. Group (Ed.), The Danielson Group. Princeton, NJ.

Darling-Hammond, L. (2003). Keeping good teachers: Why it matters, what leaders can do. Educational Leadership, 60(8), 6-13.

Dawson, M., & Lignugaris/Kraft, B. (2013). TLE TeachLivE™ vs. role-play: Comparative effects on special educators' acquisition of basic teaching skills. In <blinded for review>. Conference Proceedings for First National TeachLivE Conference. Paper presented at First National TeachLivE Conference: <blinded for review>.

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

Dieker, L. A., Straub, C., Hughes, C. Hynes, M. C., & Hardin, S. E. (2014). Learning from virtual students. Educational Leadership, 71(8), 54-58.

Earley, P., & Porritt, V. (2014). Evaluating the impact of professional development: the need for a student-focused approach. Professional development in education, 40(1), 112-129.

Elford, M., James, S., & Haynes-Smith, H. (2013). Literacy instruction for pre-service educators in virtual learning environments. In Hayes, A., Hardin, S., Dieker, L., Hynes, M., & Hughes, C., & Straub, C. (Eds). Conference Proceedings for First National TeachLivE Conference. Paper presented at First National TeachLivE Conference: Orlando, FL: University of Central Florida.

Forzani, F. (2014) Understanding "Core practices" and "practiced-based" teacher education: Learning from the past. Journal of Teacher Education, 65(4), 357-368

Guskey, T., & Sparks, D. (2002). Linking professional development to improvements in student learning. Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. Retrieved from http://files.eric.ed.gov/fulltext/ED464112.pdf

Guskey, T., & Yoon, K. S. (2009). What works in professional development? Phi Delta Kappan, 90(7), 495–500.

Hattie, J. (2003). Teachers make a difference: What is the research evidence? Camberwell, Victoria: Australian Council for Educational Research.

Hayes, A. T. (2015). The experience of presence and social presence in a virtual learning environment as impacted by the affordance of movement enabled by motion tracking. Unpublished Ph.D. Dissertation, University of Central Florida.

Hayes, A., Hardin, S., Dieker, L., Hynes, M., & Hughes, C., & Straub, C. (2013). Conference Proceedings for First National TeachLivE Conference. Paper presented at First National TeachLivE Conference: Orlando, FL: University of Central Florida.

Kane, T. J., & Staiger, D.O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. (Working Paper No. 14607). Retrieved from National Bureau of Economic Research website: http://www.nber.org/papers/w14607

Kane, T. J., & Staiger, D.O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Seattle, WA: Bill & Melinda Gates Foundation.

Lauer, P. A., Christopher, D. E., Firpo-Triplett, R., & Buchting, F. (2014). The impact of short-term professional development on participant outcomes: a review of the literature. Professional development in education, 40(2), 207-227.

Levine, T., & Hullett, C. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. Human Communication Research, 28(4), 612-625.

Loewenberg Ball, D. & Forzani, F. (2010). Teaching skillful teaching. Educational Leadership, 68(4), 40-45.

Loewenberg Ball, D., Sleep, L., Boerst, T. A., & Bass, H. (2009). Combining the development of practice and the practice of development in teacher education. The Elementary School Journal, 109(5), 458–474. doi:10.1086/599340

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. School Science and Mathematics, 99(5), 258-271.

Mathematics Assessment Resource Service. (2012). Solving linear equations in one variable [Lesson plan]. Retrieved from http://map.mathshell.org/ materials/download.php?fileid=1286

Measures of Effective Teaching Project. (2010). Danielson's Framework for Teaching for classroom observations. Retrieved from http://metproject.org/resources/ Danielson%20FFT_10_29_10.pdf

National Center for Education Statistics. (2013). NAEP questions tool. Retrieved from http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=mathematics.

National Association for Professional Development Schools. What it means to be a professional development school. Retrieved from http://napds.org/nine-essentials/

Pallant, J. (2007). SPSS Survival Manual. New York: McGraw-Hill Education.

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. Educational and Psychological Measurement, 64(6), 916-924.

Straub, C., Dieker, L., Hynes, M., & Hughes, C. (Eds). (2014). Proceedings from TeachLivE 2014: 2nd National TLE TeachLivE Conference. Orlando, FL: University of Central Florida.

Teaching Works (2014). High Leverage Practices. Retrieved from: http:// www.teachingworks.org/ work-of-teaching/high-leverage-practices#sthash. I4xK7DG4.dpuf

Thompson, M., & Wiliam, D. (2008). Tight but loose: A conceptual framework for scaling up school reforms. In E.C. Wiley (Ed.), Tight but Loose: Scaling up Teacher Professional Development in Diverse Contexts. ETS: Princeton, NJ.

U.S. Department of Education. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. Washington, D.C. Retrieved from http://www2.ed.gov/ about/bdscomm/list/mathpanel/report/final-report.pdf

Vince Garland, K., Vasquez, E., & Pearl, C. (2012). Efficacy of individualized coaching in a virtual classroom for increasing teachers' fidelity of implementation of discrete trial teaching. Education and Training in Autism and Developmental Disabilities, 47(4), 502-515.

Yoon, K. S., Duncan, T., Wen-Yu Lee, S., Scarloss, B., & Shapley, K. L. (2007). Reviewing the evidence on how teacher professional development affects student achievement. (No. Rel 2007-No. 033) (p. 55). Washington, D.C.: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel_2007033.pdf

*Lisa Dieker, Ph.D., is a Pegasus Professor and Lockheed Martin Eminent Scholar at the University of Central Florida. Her research focuses on teacher education, special education and innovations in technology.*

*Charles E. Hughes, Ph.D., is Pegasus Professor of Computer Science and co-director of the Synthetic Reality Laboratory at the University of Central Florida. His research focuses on virtual environment technologies.*

*Michael C. Hynes, Ph.D., is Pegasus Professor and Director of the School of Teaching, Learning and Leadership at the University of Central Florida. His research focuses on virtual learning environments.*

*Carrie Straub, Ph.D. is Executive Director, Educational Programs and Research at Mursion, Inc. She served as the past research director of TeachLivE at the University of Central Florida.*