Proceedings from Ludic Convergence

The Second National TLE TeachLivE™ Conference
May, 2014, Orlando, FL,
University of Central Florida
Carrie Straub, Ph.D.
Lisa Dieker, Ph.D.
Michael Hynes, Ph.D.
Charles Hughes, Ph.D.
# TABLE OF CONTENTS

- **Foreword by Drs. Lisa Dieker, Michael Hynes, Charles Hughes** .......................................................... 7

- **Second Annual TeachLivE Conference Proceedings ................................................................. 8**

- **TeachLivE for Instructional Coaches: Uses and Takeaways from District and Trainer Perspectives**
  Cherie Behrens and Janice Franceschi
  Introduction ................................................................. 8
  Implementation Procedures and Results ................................................................. 8

- **How TeachLivE is Helping Kennedy Krieger Build Model Classrooms**
  Linda Brandenburg, Claire Donehower, Deidre Rabuck
  Abstract ................................................................. 11
  Introduction ................................................................. 11
  Methods ................................................................. 12
  Results ................................................................. 12
  Discussion ................................................................. 14

- **M. Ed. in Educational Leadership Practice & Coaching in TeachLivE: Preliminary Findings**
  Hilary Buckridge and Rosemarye Taylor
  Introduction ................................................................. 16
  Purpose of the Study ................................................................. 16
  Procedures ................................................................. 17
  Data Analysis ................................................................. 17
  Next Steps ................................................................. 18

- **The Influence of TeachLivE on Anxiety Levels in Preservice and Inservice Mathematics Teachers**
  Heidi Eisenreich and Katie Harshman
  Abstract ................................................................. 19
  Review of the Literature ................................................................. 19
  Statement of the Problem and Purpose of the Study ................................................................. 20
  Method ................................................................. 20

- **Measuring Social Presence and Impact of Social Presence in TeachLivE: Ludic Convergence**
  Aleshia Hayes ................................................................. 23

- **Trial-Based Functional Analysis in Virtual Environments for Teacher Preparation**
  Aaron Koch, Eleazar Vasquez, Matthew Marino, Carrie Straub, Kelly Schaffer, Claire Donehower
  Introduction ................................................................. 27
  Method ................................................................. 28
  Results and Discussion ................................................................. 28

- **English Learner (EL) Avatars for Pre-Service Teachers**
  Michele Regalla, Joyce Nutta, Cynthia Hutchinson, Pam Jones, Nooshan Ashtari, Karen Verkler
  Introduction ................................................................. 30
  Steps and Facilitation ................................................................. 31
  The Courses: TSL 4080 and EDG 4410 ................................................................. 32
  Procedures ................................................................. 32
Foreword

Who knew? When the three of us started this journey and began our friendship, we never dreamed it would expand beyond UCF, and yet, today, we have over 50 partners across the globe. We are shocked by the ludic convergence we continue to experience daily with our UCF family and the true stars of our work: Maria, CJ, Sean, Kevin, and Ed. One of the best parts of our work is each of you. The ability to talk, play, laugh and come together at the 2nd Annual TLE TeachLivE™ Conference in Orlando was once again a success beyond what any of us had ever dreamed. The highlight of the conference was the fun kickoff provided by the amazing and generous Cheryl Hines, a proud UCF alum. The style and substance of her presentation reminded us that, if things are not ludic, why bother. She also challenged us as educators to use improvisation skills in our preparation of great teachers and leaders.

Again this year, we humbly write the foreword but the true credit for the conference goes to the work of Taylor Bousfield and Benjamin Gallegos, two doctoral students in special education, who led the conference planning. We would be remiss not to acknowledge numerous other folks at UCF, but if we listed them all we would surely miss someone and the entire foreword would be a list of names. In any case, we hope what you learned from your attendance is that we are great hosts who know how to eat, have fun and, most importantly, meet our goal as a university to be the leading partnership university in America. Having each of you and your sites as current, future and/or continuing partners is a privilege that has allowed us to learn continually from each of you. We hope the over 130 conference participants truly experienced ludic convergence during the wonderful keynote session by Scott Benson from the Bill & Melinda Gates Foundation, through the fantastic sessions provided by amazing colleagues, and from the always forward-thinking research and the great candy bar at the SREAL Lab. We also acknowledge our fantastic program officer, Amy Slamp, from the Bill & Melinda Gates Foundation. That funding allowed our kids to grow up and to share our research outcomes from our first year national study.

We continue to believe that the true spirit of TLE TeachLivE™ is not in the technology nor in the avatars, but that the real spirit of TLE TeachLivE™ is in the ludic convergence we all continue to have playing in the sandbox with the avatars to change the landscape of teacher preparation. We know there are still so many lessons to be learned and so many more products, scenarios and actions to be developed. We hope you enjoy reading the abstracts of the great folks who attended and submitted their work to be included in this proceeding document from the 2nd Annual TLE TeachLivE™ Conference. The articles submitted show the depth and ludic convergence of our partners. Thanks for being such amazing and “wicked smart” people. We hope you all continue to take risks with us to think differently, always remembering the definition of ludic convergence: playfully merging these diverse areas of expertise into a network of partnerships that stimulates innovation; providing a fertile ground for new collaborations that engage all stakeholders and partners.

Lisa Dieker, Michael Hynes, and Charles Hughes
University of Central Florida
TeachLivE for Instructional Coaches: Uses and Takeaways from District and Trainer Perspectives

Cherie Behrens  
University of Central Florida  

Janice C. Franceschi  
The School District of Osceola County, Florida

Introduction

Guiding research indicates that for instructional coaches to positively affect the teaching practices of teachers, coaches need professional development to ensure they are more knowledgeable and skilled than the teachers they are coaching (Dole, 2004; Marsh et al., 2008). Thus, Janice Franceschi, the Director of Professional Development, and Assistant Superintendent Lissette Brizendine, of the School District of Osceola County met to discuss the kind of professional development desired based on the results of their district coaches' needs assessment and national, federal, Race to the Top, state and district priorities. Priority examples entail the Learning Forward Standards and Florida's State Board of Education Strategic Plan Goals, which include concentrated efforts in promoting student achievement and long-term student success. Considering that a growing body of research indicates that coaches can significantly and positively impact student achievement through the teachers they coach, Franceschi and Brizendine began planning tailored professional development based on the outlined needs (Bean et al., 2010). They consulted Cherie Behrens, a University of Central Florida lecturer with a history of impacting student success at district and state levels. Cherie's work drew upon her creation of the first state-approved, graduate-level, instructional coaching course combined with her district to school level coaching experiences to develop a several-day, needs-tailored, Instructional Coaching Series.

The three full-day, train-the-trainer professional development offerings were held for district coaches this academic year on February 10th, March 10th and April 7th from 8:30 am to 3:30 pm each day; however, a few site-based coaches also attended to take advantage of extra space available. In order for the attending coaches to receive in-service points, they were required to attend all three training days and develop an implementation plan. The formats of the training were whole group, small group, and one-on-one, with differentiated instruction provided based on in-the-moment formative assessment measures.

A major component of the Instructional Coaching Series involved the use of an innovative, Bill & Melinda Gates-funded technology called TLE TeachLivE™ (TeachLivE). TeachLivE was the most impactful technology utilized during the sessions according to data collected from the coaches. TeachLivE is a classroom simulator used to better train educators. The TeachLivE simulator is populated with specifically designed avatars that interact with educators. Using differentiated learning techniques and a formative assessment feedback model, TeachLivE provides each user or user groups with a different experience that changes based on the actions of the educator(s) in the simulation.

Implementation Procedures and Results

For every instructional coaching session, TeachLivE provided coaches with a center to further hone their abilities coaching a resistant teacher (an area in which school-based coaches, in a December 2013 survey, reported to struggle), practicing what was learned in the sessions,
utilizing "look for" documents for further coaching conversations and more in an unpredictable but realistic way, using a mixed-reality environment. Research suggests that effective coaches must provide a specialized kind of feedback to teachers, feedback that involves non-judgmental and reflection-inducing comments or questions based on data collected watching a teacher teach and listening to the comments a teacher makes (Knight, 2011). This specialized feedback, referred to as "coaching feedback", aims to cause and promote teacher reflection; evoke positive changes in teaching behaviors, pedagogical and content knowledge; and impact overall teaching practices. Thus, time in TeachLivE enabled teachers to strengthen their coaching feedback skills and overall coaching abilities.

Formative data collected after the first session in TeachLivE demonstrated that many of the coaches needed more practice coaching a resistant teacher. So, instead of choosing a different simulation in TeachLivE for the next session, coaches received another opportunity to practice coaching the resistant teacher with more supportive measures, including a model by Cherie on how to coach the resistant teacher using the language templates the coaches were given and encouraged to use. This kind of flexibility allowed the group to better meet its mastery goals. After Cherie modeled coaching, the entire group talked about the session and Cherie gave them more resources about success with resistant teachers. After the coaches had opportunities to coach the resistant teacher, Cherie alone or with the coach's PLC (if they were comfortable with the group watching) would provide feedback and coach the participant to stronger coaching. Here are data from coaches about this particular experience: TeachLivE—got so much more this time around! Loved the model coaching done at the beginning and the coaching at the end. Very reflective. Cherie's comments were very helpful and useful”—Debbie; "Great experience, must utilize more coaching language, it's a great experience”—No name; "Being in the hot seat was a challenge, but what challenges you grows you! The feedback was immediate and valuable and I am so glad that I had the opportunity to practice. Having my peers in there with me also was encouraging and I believe we all learned in this process”—Joyce.

A snapshot of TeachLivE activities over the course of the Instructional Coaching Series:

Day One: Coaches used their resources (including articles that provided guidance for successful coaching with resistant teachers), plus coaching language stems, to coach a resistant teacher* named Ms. Adkins in TeachLivE. Ms. Adkins used the identical language reported by coaches in the December 2013 survey to replicate, as closely as possible, what coaches are hearing at schools from resistant teachers. Cherie also drew upon her experiences to inform Ms. Adkins’ language, in order to enable what philosopher Samuel Taylor Coleridge refers to as “a suspension of disbelief.”

*Resistant teachers may resist change, coaching, school initiatives, etc.

Day Two: Cherie modeled coaching Ms. Adkins, the resistant teacher, in the whole group meeting first and then, in rotations, coaches individually coached Ms. Adkins while Cherie observed. Afterwards, Cherie provided feedback to optimize their coaching alone or with additional input from the individual's Professional Learning Community (PLC) who watched the coaching occur.

Day Three: The coaches watched Cherie act as a struggling teacher and marked observations on a “look-for” document to have data-focused coaching conversations later on.
Afterwards, Cherie and the instructional coaches discussed how to coach the teacher, triage the teacher’s needs so as to not overwhelm the teacher, etc. In their following center, the instructional coaches role-played with each other as if they were coaching the struggling teacher, Cherie, they had observed in TeachLivE.

Overall, TeachLivE enabled coaches with opportunities to practice and strengthen their coaching skills in closely realistic settings. Thus, TeachLivE was an excellent pairing for this professional development series. After receiving such positive results using TeachLivE, the School District of Osceola County became a partner with TeachLivE and have since installed TeachLivE at their professional development headquarters for further use.

References


How TeachLivE is Helping Kennedy Krieger Build Model Classrooms

Linda Brandenburg  Claire Donehower  Deidre Rabuck
Kennedy Krieger Institute

Abstract

Kennedy Krieger School Programs are part of the Special Education department of Kennedy Krieger Institute, a research and rehabilitation hospital affiliated with Johns Hopkins Medical Institutions in Baltimore, Maryland. The school programs include a K-8 program, a high school program, and two specialized programs for students with autism spectrum disorders (ASD). The Kennedy Krieger School Programs began implementing the Model Classroom Program to accomplish two goals: first, to assess a teacher’s implementation of best practices in the classroom, providing individualized professional development through ongoing feedback and coaching in the areas collaboratively targeted by the teacher and the coach for improvement; second, to evaluate the aggregate of individual teacher performance in order to better determine the need for programmatic initiatives in the school. Ongoing professional development is essential for keeping educators informed and helping them feel confident as they gain experience in the classroom. The goal of this project was to add TeachLivE to the coaching component of the Model Classroom Program. Providing teachers with the opportunity to practice target behaviors in the mixed-reality simulator, receive immediate coaching and feedback, and then practice again within the same lab session, proved to be a valuable tool in facilitating their acquisition of the target behaviors.

Introduction: Context, Research Problem, Review of Literature

There are several core concepts or beliefs that support the implementation of the Model Classroom Program. These include the belief that teacher effectiveness is the key to positive outcomes for students, and that the variables used to identify teacher effectiveness must be clearly defined. Additionally, teacher effectiveness continually improves in quality with supervision, evaluation, and ongoing communication and feedback to assess the effectiveness of the interventions through coaching (Fixen et al., 2005). Such a program of improvement of teacher effectiveness will be labor-intensive, requiring real commitment from the teachers involved, the coaches, supervisors and administrators (Reinke, Lewis-Palmer, & Merrell, 2008).

The Model Classroom is composed of elements that lead to student success in learning. The environment for effective instruction must be structured with a physical setup that keeps the students engaged and comfortable, and provide a comprehensive, understandable schedule, with visual supports as needed (McGill et al., 2003). Lesson planning must relate to meaningful curriculum, and lessons must be delivered by a well-managed staff and with appropriate collaboration with paraeducators and related service providers. Effective teacher behaviors align with effective classroom and behavior management, and include the use of an ongoing data collection system with clearly identified targets to be evaluated (Espin & Yell, 1994).

In order for research-based practices to be successfully implemented into classroom settings, the components of implementation science must be present: exploration, installation, initial implementation, full implementation (Odom, Cox, & Brock, 2013). The smooth flow of teacher behaviors is essential, and the coaching model (i.e., installation and implementation) provides the teacher with an assessment as well as feedback about their ability to implement best
practices. Using a repeated, structured measure to evaluate the teacher, the program includes individualizing target interventions and striking a balance between the fidelity of implementation and the flexibility needed for each teacher, students, and classroom environment. The feedback is ongoing and progress is monitored continuously, with appropriate reinforcement or reward provided to the teachers as they demonstrate success in implementing the program.

The goal of this project was to utilize the TeachLivE lab to allow teachers to practice some of the target behaviors identified in the Model Classroom Program. Target behaviors included: opportunities to respond (Sutherland & Wehby, 2001), error correction (Barbetta, Heron, & Heward, 1993), and praise rates (Cameron & Pierce, 1994), including the ratio of specific to general praise (Musti-Rao & Haydon, 2011). Teachers were observed in their classrooms, given feedback and coaching, participated in scheduled lab time for teaching the avatars, got immediate feedback from the coach, and then had the opportunity to teach again within the same lab session.

Methods

Fifteen teachers were selected from our four school programs to participate in TeachLivE lab sessions. The teachers had a range of teaching experience from one year to 10 plus years in the classroom. During each session, two to three teachers and at least one coach (e.g., administrator, supervisor, etc.) were present. The participating teachers were provided with a lesson plan and lesson materials prior to the TeachLivE session. Teachers who were new to the lab were given a brief demonstration of how to interact with the students and move around the classroom. The format of the sessions was consistent from one session to the next.

1. Teacher 1 would teach for five to seven minutes.
2. Teacher 1 would receive feedback from coaches and other teachers.
3. Teacher 2 would teach for five to seven minutes.
4. Teacher 2 would receive feedback from coaches and other teachers.
5. Teacher 3 (if applicable) would teach for five to seven minutes.
6. Teacher 3 (if applicable) would receive feedback from coaches and other teachers.
7. Each teacher would have a chance to apply the feedback and teach a second time followed by brief feedback from coaches and other teachers.

During each teaching trial, quantitative data was collected on the teacher’s pacing (providing students with opportunities to respond), praise rates, praise specificity, and error correction. Qualitative data was collected on other teaching variables (e.g., content knowledge, building rapport with students, managing off task behavior, etc.). For many teachers, this data was collected in their classroom prior to the lab session(s) and following the lab session(s) to determine whether or not the target behaviors were generalized into the classroom environment.

Results

In general, there were positive results both within the TeachLivE sessions (between teaching trials) and from TeachLivE into the classroom. The examples below illustrate some of the positive results from the first year of this partnership.

Figures 1 and 2 show some of the positive effects within TeachLivE sessions. In both of these examples, the teachers were able to show an improvement in one of their target behaviors from their first teaching trial to their second teaching trial. Figure 1 shows that a teacher was able
to improve her pacing (opportunities to respond) from 2.0 per minute to 3.05 per minute on average from the first trial to the second trial within a lab session. Figure 2 shows that a different teacher was able to improve her praise rates from 1.9 per minute to 2.8 per minute on average from the first trial to the second trial within a lab session.

![Teacher CS](image1)

**Figure 1.** The average scores for opportunities to respond (OTR) for Teacher CS from the first to second trials within a lab session.

![Teacher SS](image2)

**Figure 2.** The average scores for rate of praise for Teacher SS from the first to second trials within a lab session.

Figures 3 and 4 show the generalization of target skills into the classroom environment. In both of these examples, the teachers were able to show an improvement in one or more target behaviors not only within the lab sessions, but also in their classrooms. Figure 3 shows a teacher who was able to improve her lesson pacing (opportunities to respond) and increase her rates of
praise within her classroom after only one TeachLivE session. Figure 4 shows a teacher who required two sessions in the lab targeting her praise specificity before she was able to generalize that skill into the classroom.

**Figure 3.** The rates per minute for opportunities to respond (OTR) and praise for Teacher RJ across lab sessions and classroom samples.

**Figure 4.** The percent of specific praise vs. general praise for Teacher JE across lab sessions and classroom samples.

**Discussion**

Teachers showed improvements in their target areas between their trials in the lab, and many were able to generalize what they learned back into the classrooms. Teachers also enjoyed being in the lab after their initial nerves settled down. One teacher noted, “Honestly, it might be
the most beneficial teacher education tool I have seen, with the instant feedback and re-teaching opportunities. Universities should invest, and use it with undergrad students. Also, I think my specific praise dramatically increased today! Thanks for telling me about it and signing me up!"

Others felt that the experience helped their teaching skills and gave them an opportunity to meet other teachers, get immediate feedback, hear the feedback given to other teachers, etc.

Following the lab sessions, the teachers and coaches involved in this project gathered to discuss the pros and cons of their experiences and brainstorm next steps or future directions for TeachLivE and the Kennedy Krieger School Programs. Some of the ideas discussed were to: practice asking questions that access higher order thinking skills; practice with discrete trial teaching using an avatar with autism; prepping for IEP meetings or parent conferences with an adult avatar; training clinicians on a specific intervention or strategy with students or parents.

References


M. Ed. in Educational Leadership Practice & Coaching in TeachLivE: Preliminary Findings

Hilary Buckridge
Orange County Public Schools, Florida

Rosemarye Taylor
University of Central Florida

Introduction

Utilizing multimedia methods and virtual environments are now part of educational preparation programs for surgical medical programs, flight simulators, and military combat training situations, and believed by researchers to become a part of the educator preparation programs, meeting the needs of the Generation M population in academic settings (Dieker et al., 2012). TeachLivE virtual education simulation technology, originally designed for teacher education simulation with student avatars, has expanded to include educational leadership simulations utilizing adult parent and teacher avatars.

The M. Ed. in Educational Leadership at the University of Central Florida began incorporating the mixed reality resource of TeachLivE during the fall semester of 2013. Two experiences are provided. One is a parent conference and one is a post-observation conference with a teacher. Educational Leadership master’s degree students need preparation and practice with feedback in communicating with parents and in providing post-observation feedback to teachers in an administrative capacity before entering the administrative internship. By maximizing mixed reality technology of TeachLivE with side-by-side coaching, practice situations of administrative-level parent conferencing and teacher conferencing are available. Feedback from this practice is intended to scaffold students to the administrative internship experience. This power of the sequencing of instruction through scaffolding is the guided and independent practice model, using realistic scenarios and simulation practice with students (Taylor, 2010).

Purpose of the Study

The purpose of the study is to ascertain the effectiveness of mixed reality experiences using TeachLivE for Educational Leadership M. Ed. students. The use of avatars and virtual teaching provides authentic practice where mistakes do not impact real students, and through reflective practice, feedback and coaching, novice educators can deepen their practice (Dieker et al., 2008).

The second purpose is to determine the perceived value of coaching feedback received immediately following mixed reality experiences and whether the experiences were perceived as valuable later in the M. Ed. program after completion of the administrative internship. Leadership through coaching in Hersey’s situational leadership model, cited by Bolman and Deal (2008) is effective in high relationship and high task situations with followers willing and motivated to learn.

Research questions addressed in these proceedings follow.

1. To what extent do Educational Leadership M. Ed. students believe the TeachLivE parent conference and teacher post observation conference simulation experiences to be helpful in developing their communications skills with parents and teachers?

2. To what extent do Educational Leadership M. Ed. students believe the post observation feedback was helpful in developing their communications skills with parents and teachers?
**Procedures**

Scenarios are written in which the interactor (who directs the avatars) and the students have reviewed in preparation for the experiences. For each of the two experiences, students receive immediate coaching from an expert who gives supportive but direct feedback on the experience. Students then write a reflection on the experience and submit to their instructor. Coaching is a shaping of behavior by observing performance, offering guidance, and recommending specific practice to emphasize (Owens & Valesky, 2011).

The researcher is collecting qualitative data documenting students’ perception of the value of the TeachLivE experience and the coaching feedback to provide authentic virtual rehearsals as a future school administrator in the two experiences: communicating with parents and teachers through conferencing. Reflections on the experience will also be analyzed for themes and commonalities in 2015 and are not addressed in these proceedings.

Data collection began in the fall semester 2013 with the M. Ed. Educational Leadership students enrolled in the face-to-face courses only, coinciding with the Supervision and Parent & Community coursework. The mixed reality experience is part of the scaffolded instruction process taking theoretical research-based knowledge, and providing specific targeted skills practice before entering live situations during internship. Students receive an orientation to TeachLivE during class, and receive a copy of each of the four possible scenarios. Students sign up in pairs for 30-minute blocks in the simulation lab. Each student receives 10 minutes of TeachLivE interaction followed by five minutes of immediate coaching. Next, the student observes the partner’s session.

Faculty and/or an expert coach provide just-in-time feedback/coaching at the conclusion of the 10-minute interaction. This practice aligns with the review of over 8,000 studies led by Hattie (1992) as cited in Marzano (2003) to indicate that the most powerful single modification that enhances student performance and deep learning is feedback. To be most impactful, feedback must have two characteristics: it must be timely and specific (Marzano 2003); these characteristics are employed with these students.

After completion of a two-semester administrative internship, and practice with teachers and parents, the M. Ed. Educational Leadership students will complete an exit survey. Two items related to the impact of TeachLivE have been added and results will be analyzed.

**Data Analysis**

Initial data was gathered from the fall semester of 2013 Supervision course, Post Teacher Observation Conference $n=13$ and the spring semester of 2014 Community and School course, Parent Conference $n=17$. The feedback form had five items with a Likert scale with values from one to five with five being the most positive. The items relate to how beneficial the experience was, whether the experience should be continued in the M. Ed. Program, perception of improving speaking confidence, helpfulness of immediate feedback, and the extent to which the experience was realistic. An open-ended item was also completed but is not a part of this preliminary analysis.

Thirty students from both semesters completed feedback forms: 17 in the parent conference and 13 in the post-observation teacher conference. Students’ responses indicate both experiences were helpful in preparing them and improving communication. Feedback from the knowledgeable coach is ranked highest among the items. These findings can be reviewed in Table 1.
Table 1. Fall 2013 and Spring 2014 on 5-point Likert Scale, five most positive ranking, \( n = 30 \).

<table>
<thead>
<tr>
<th>Survey</th>
<th>Item</th>
<th>Mean</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Conference</td>
<td>Beneficial</td>
<td>4.76</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Simulation Continuation</td>
<td>4.82</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Speaking Confidence</td>
<td>4.41</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Feedback Helpful</td>
<td>4.88</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Realistic</td>
<td>4.56</td>
<td>17</td>
</tr>
<tr>
<td>Post Observation</td>
<td>Beneficial</td>
<td>4.92</td>
<td>13</td>
</tr>
<tr>
<td>Conference</td>
<td>Simulation Continuation</td>
<td>4.92</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Speaking Confidence</td>
<td>4.46</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Feedback Helpful</td>
<td>5.00</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Realistic</td>
<td>4.69</td>
<td>13</td>
</tr>
</tbody>
</table>

**Next Steps**

Data will continue to be gathered through spring semester 2015. At that time all feedback will be analyzed on the quantitative items by perceived value and by demographic factors of the participants. Qualitative open-ended item responses and the students’ reflections will also be analyzed. The study is anticipated for completion by summer 2016.

**References**


The Influence of TeachLivE on Anxiety Levels in Preservice and Inservice Mathematics Teachers

Heidi Eisenreich    Katie Harshman
University of Central Florida

Abstract

Doctoral students conducted a research project in Spring 2014 using a galvanic skin response device to measure anxiety levels of preservice and inservice teachers during interaction with the TeachLivE students. Preliminary results will be discussed.

Review of the Literature

School districts are criticizing teacher preparation programs with claims that new teachers are not prepared to enter the classroom. However, colleges and universities are not always able to place prospective teachers in internships due to accountability requirements for teachers. Virtual classroom environments could aid in teacher preparation. In these environments, preservice teachers are able to practice playing the role of student or teacher to better prepare them to enter the classroom. Programs such as TeachLivE offer a mixed-reality experience in which teachers are able to deliver instruction and manage behavior without the worry of putting children at risk. These types of environments also offer the added benefit of immediate feedback from the instructor, and the ability to revise any issues and reteach as many times as needed.

TeachLivE, developed at the University of Central Florida, is a virtual classroom that provides preservice and inservice teachers with an environment in which to practice their pedagogy without placing real students at risk (About TeachLivE, 2013). The participant faces a large screen and stands in front of student avatars that are controlled by a specifically trained actor in a remote location. Depending on the objectives of the experience, the actor impersonates typically developing or not typically developing students. Participants are able to present new content, review content, practice behavior management skills, practice a specific technique such as scaffolding, and perform many other daily tasks that a teacher would experience. If a teacher is not successful during a session, he or she may go back and practice repeatedly, without risking the education of a real student. A teacher may step away from the screen to face the entire class, or walk forward and lean in to interact with individual students (Andreasen & Haciomeroglu, 2009).

Doctoral students conducted an experimental study to determine the influence of TeachLivE on preservice and inservice mathematics teachers’ anxiety levels using the Zung Anxiety Self-Assessment Scale (Zung, 1971) and a Galvanic Skin Response (GSR) device.

The Shimmer GSR is a wireless device that uses two electrodes, which are placed on the fingers to measure the user’s skin conductance to monitor stress levels (Handri, Nomura, Kurosawa, Yajima, Ogawa, & Fukumura, 2010). The device sends data to a coordinator via a wireless technology called ZigBee, and the coordinator sends information to a computer. From there, data can be analyzed using a variety of applications (Villarejo, Zapirain, & Zorrilla, 2012).

A major advantage to virtual classrooms such as TeachLivE is the immediate feedback participants are able to receive. “Feedback is essential to learning, and recent research suggests that the most effective feedback is immediate rather than delayed” (Scheeler, McKinnon, & Stout,
In conventional programs, teacher interns will video record a lesson that they teach in their assigned classroom and submit it for review at the intern’s next class meeting. It could be a week or more before feedback is provided. Virtual classrooms alone provide participants with the essential feedback, as the professor is typically in the room or observing via the Internet. Additionally, these programs can be coupled with other technologies, such as bug-in-ear, which could provide even better teacher preparedness.

Statement of the Problem and Purpose of the Study

Colleges and universities have difficulty placing prospective teachers in internships due to accountability requirements for teachers. The literature supports the effectiveness of virtual classrooms such as TeachLivE. The purpose of this study is to gain information regarding anxiety levels of preservice and inservice teachers in TeachLivE—specifically, to determine if anxiety levels decrease with extended time in TeachLivE according to the GSR device, and if the post scores from the student-responded anxiety survey instrument vary among students who play video games at different levels (never, rarely, sometimes, and frequently). Since the research on TeachLivE is limited, we will be adding to a small pool of articles on TeachLivE. The majority of articles surrounding TeachLivE focus on the inner workings of the system. The few research studies obtained utilized TeachLivE in special education and teacher preparation. As this is a first experiment, there are no previous findings. However, researchers expect to find that anxiety levels decrease after preservice and inservice teachers interact with the TeachLivE student-avatars.

Method

Research Questions
1. Do anxiety levels decrease with extended time in TeachLivE according to the GSR device?
2. Do post scores from the student-responded anxiety survey instrument vary among students who play video games at different levels (never, rarely, sometimes, and frequently)?

Research Design

This study utilizes a one-group pretest-posttest quasi-experimental research design (Gall, Gall, & Borg, 2007), with volunteer research participants. Before collecting data, the researchers obtained approval from the institutional review board (IRB), and requested permission from course leaders to use their students. After discussing the research study, the course leaders allowed their students to replace one of their assignments with participation in our research. The researchers then went to the classes to discuss the research study with the preservice and inservice teachers, explained confidentiality, and made it clear that participants could opt out of the study at any time.

Population and Sampling

The researchers employed convenience sampling, choosing readily available participants (Gall et al., 2007) who took a pretest and posttest and were not randomly. The treatment group included 23 Mathematics Education undergraduate and graduate students at the University of Central Florida, enrolled in a math methods course; 3 were undergraduate and 20 were graduate students; 8 were male and 15 were female, and 19 had no teaching experience prior to this school year. Leaders for the methods course were doctorate level mathematics educators at the University of Central Florida, and researchers were completing this research as part of the doctoral program.
Data Collection Procedures

In order to determine how TeachLivE affected anxiety levels of research participants, the researchers: (i) administered a pretest; (ii) administered the treatment; (iii) administered a posttest. The pretest was administered immediately preceding the treatment. The treatment was administered immediately after the pretest, and the posttest was administered immediately after the treatment. All three steps were completed within one day.

Instrument

The instrument chosen for this study was the Zung Anxiety Self-Assessment Scale. Students completed this before and after the TeachLivE simulation. Researchers added up the raw scores given by authors of the Zung Anxiety Scale with the following interpretation: below 45 = within normal range, 45-59 = minimal to moderate anxiety, 60-74 = severe anxiety, 75 and over = most extreme anxiety.

Researchers also placed a GSR device on the participants for two minutes prior to interaction in the simulation to establish a baseline. Participants then interacted with the TeachLivE students for five minutes and data was collected wirelessly via the GSR. Researchers chose the GSR Device because it “detects whether there has been an effort or a different situation from being relaxed with a success rate of 90.97%” (Villarejo et al., 2012, p. 6099).

Data Analysis/Statistical Procedures

The researchers conducted an analysis of variance (ANOVA) for repeated measures to determine whether the anxiety levels, as measured by the GSR, decreased with extended time in the TeachLivE simulation (Gall et al., 2007). Since Mauchly’s test of Sphericity was violated we used Greenhouse-Geisser results. It was significant at the .05 level, indicating that over time in the simulation, students’ anxiety levels decreased. Our eta squared was .416, considered a very large effect size. This indicated that the students’ anxiety levels were affected by being in the TeachLivE simulation, in particular from minute two to minute three.

Though no statistical significance was found between the student-reported post anxiety scores in relation to the amount of time they spent playing video games, there was a difference in the means (Never: $M = 36.5$, Rarely: $M = 31.29$, Sometimes: $M = 30.25$, and Frequently: $M = 30.33$). All anxiety scores fall within the normal range of anxiety, but this data shows students that played video games more often had lower anxiety than students who played them less often.

Limitations

A possible limitation in this study lies in the research design. Quasi-experimental research suffers threats to internal validity such as statistical analysis, meaning posttest score differences are due to pre-existing group differences. Since the study utilized convenience and voluntary sampling, there may be selection bias and issues with generalization. Another limitation was the small sample size of 23, and that only 14 had useable GSR data. Further, of the 14, we were concerned with validity of the results because the GSR device may not have had enough time to fully dry in the minute or two between participants. In the future, it would be better to have multiple devices to allow the sensors time to fully dry.
Pretest and posttest questions were identical, so students already knew the questions the second time around, and this could have affected their self-reported anxiety for the posttest. Since students self-reported their anxiety levels, they could also be biased. Some of the students had a high anxiety level at the start of the simulation, as measured by the GSR device. Specifically, starting anxiety levels ranged from 300 to 2400. However, upon completion of the treatment, participants’ final readings ranged from 199 to 1300.

**Implications**

Since anxiety levels drop significantly from minute two (first minute they were in the simulation) to minute three, this could imply that students are more anxious before the simulation, and relax once they enter and begin talking to the avatars. There are many possible explanations for this: students are more anxious before the simulation because they do not know what to expect, and they feel more at ease once they get into the simulation and begin speaking with the TeachLivE avatars, etc. Students who never play video games could be more anxious in the TeachLivE simulation because they are not used to being in a simulated world.

Future research should include repetitions of the study with different preservice teacher samples, and should include a control group to determine the effect of the GSR device. Repeated findings would give stronger evidence of validity and be more generalizable (Gall et al., 2007).

**References**


Aleshia Taylor Hayes
University of Central Florida

Education technology is a growing multi-billion-dollar industry (Booker, 2013). Expansion of online and virtual platforms including MOOCs (e.g., Coursera and edX), Khan Academy, and Virtual School, is becoming pervasive in both urban and rural school districts (Allen & Seaman, 2013; Allen & Seaman, 2014). The research summarized in this paper is meant to examine the interpersonal connections in Virtual Learning Environments (VLEs) and is intended to inform practitioners in education as they analyze the benefits and costs of emerging platforms for education. TeachLivE is a mixed reality classroom that focuses on virtual rehearsal of interpersonal skills to master (Dieker, Hughes, Rodríguez, Lingnugaris-Kraft, Hynes, & Hughes, 2014; Hayes, Straub, Dieker, Hynes, & Hughes, 2013; Sawchuck, 2011).

This research has been iterated in multiple pilots and has evolved over time. This progression has culminated in establishing a need to draw a clear distinction between presence and social presence, as the focus on the learning and the experience is primarily about the social interaction. Presence, specifically physical presence, is summarized as the experience of “being there” in a virtual environment (Minsky, 1980; Sheridan, 1994; Witmer & Singer, 1998). The International Society for Presence Research has synthesized the definitions of presence to “when part or all of a person’s perception is directed toward objects, events, and/or people created by the technology, and away from objects, events, and/or people in the physical world. Note that the person’s perception is not directed toward the technology itself but the objects, events and/or people the technology creates.” Social presence, on the other hand, can be summarized as the sense of feeling connected to another consciousness, and is not necessarily contingent on physical proximity. The current research has synthesized social presence to a “sense of mutuality in which a user feels connected with another entity based on interactions that transcend the medium” (Bailenson, Aharoni, Beall, Guadagno, Dimov, & Blascovich, 2004; Burgoon & Harms, 2002; Lombard & Ditton, 2000). This is not to marginalize the research that has concluded that there is a relationship between social presence and presence, but instead to build upon it by isolating the variables that are most relevant to the topic of social learning.

This paper continues by outlining future research that can add clarity to the decisions made by educators, administrators, and policy makers. Further, the value of social presence in VLEs is being investigated. Finally, the approaches to measuring social presence are being examined. The participants of this research are practicing teachers who intend to improve on their pedagogy by immersing themselves in TeachLivE virtual rehearsals of classroom teaching. This exploration of the perceptions and learning outcomes of teachers in a VLE reveals trends that should inform decisions about VLE elements that provide the greatest benefit.

Over the course of this research, presence has become a smaller construct, while social presence has emerged as the relevant construct. Much of the research highlights the importance of physical presence for learning physical tasks (Patel, Bailenson, Jung, Diankov, & Bajcsy, 2006). The literature and the pilot both suggest a symbiotic relationship between social presence, engagement, learning of social behaviors, and presence. Initially, the Witmer and Singer (1998) presence questionnaire was administered to participants teaching in both their classroom and,
subsequently, in TeachLivE. This revealed a transfer of learning from TeachLivE to the teacher’s actual classrooms and high ratings of presence in which participants reported feeling like they are “there” in class with the virtual students (Hayes et al., 2013).

The initial findings from pilot one supported what the developers and users of the system intuitively believed, that the participants experienced presence. This, however, induced another research question: How much presence, and what does the sense of presence actually do to learning? Further, the researcher found that the data from the presence instrument was too subjective to draw meaningful conclusions about each experience in relation to another.

Further literature review revealed that the subjectivity of presence instruments was a common contention among researchers (Bailenson et al., 2004; Biocca, Harms, & Burgoon, 2003; Slater, 2004). In “How Colorful Was Your Day,” Mel Slater stated it most directly: “[R]esearchers interested in studying the concept called presence might find a way to abandon the easy but ultimately useless employment of questionnaires, and search for a better way to capture this elusive concept…There are many responses to a VE experience— gross behavioral, eye movements, measurable physiological responses, what people say in interviews, how people respond to questionnaires. Why elevate just one of these to be the preeminent position that it now has? Is it only because it is the easiest approach to take?” (2004). This review of pilot one also highlighted the absence of the critical construct, social presence (Hayes, Hardin, & Hughes, 2013; Zhao, 2003).

The second pilot included qualitative and quantitative measures for both presence and social presence as constructs. While the learning still included the social behaviors of delivering specific praise and asking higher order questions, the participants shifted from being teachers in K12 classrooms to being aspiring college professors. One of the outcomes of this second pilot was the creation of a behavioral coding sheet that reflected empirical observation and the social presence literature. The research also revealed contradictions between behavioral measures of social presence and the self-report, in which participants who behaved as if they were engaged and connected to the virtual students reported that they did not feel connected or engaged.

The inconsistencies between the behavioral measure and the subjective measure demanded the inclusion of a more objective measure. Consistent with the literature, the objective measure that will be added to the study is physiological data, by way of participant heart rate (Meehan, Insko, Whitton, & Brooks, 2002; Meehan, Razzaque, Insko, Whitton, & Brooks, 2005). Heart rate was determined to be the most appropriate for this research as it is an established measure for stress, engagement, and social presence. The behavioral measures include a tally of behaviors that literature and empirical research indicate social presence while the participant is teaching in the virtual classroom (Bailenson, Yee, Merget, & Schroeder, 2006; Serby, 2011). Similarly, the physiological measures are taken during the participants’ virtual rehearsal. Finally, the subjective measures will be the social presence instrument and the presence instrument that participants complete after the immersive experience.

The final step of this leg of the research will be to integrate physiological, behavioral, and subjective measures of presence. The intent of this synthesis will be to investigate any potential correlations between these different measures. Additional research in this area will explore the implications of physical changes to the experience of presence and social presence. This may yield analysis of the constructs’ possible mediating or moderating effects on one another as well as on learning outcomes.
References


Introduction

A functional analysis is a controlled procedure for evaluating behavior function. The completion of an analog condition functional analysis requires the systematic manipulation of controlled environmental variables (Carr & Durand, 1985; Iwata, Dorsey, Slifer, Bauman, & Richman, 1982/1994). These manipulations are completed in order to precisely identify the maintaining variable or variables associated with a specific target behavior (e.g., aggression, stereotypy). By executing a series of analog conditions, investigators are able to track relevant reinforcement contingencies and formulate a functional hypothesis of maintaining variables (i.e., situations/factors associated with the problem behavior). These contingencies are considered representative of those occurring in an integrated fashion within the natural environment.

Three research papers have shown that lay individuals can implement functional analysis with minimal training. Iwata et al. (2000) showed that upper-level undergraduate students could be trained to implement the attention, control, and demand conditions of a functional analysis. The study has been systematically replicated by Moore et al. (2002) and Wallace, Doney, Mintz-Resudek, and Tarbox (2004). Moore et al. showed three teachers trained to implement the demand and attention conditions of a functional analysis. Wallace et al. showed two teachers and a school psychologist trained to implement attention, control, and demand conditions.

All three studies used a multiple-baseline, across-subjects design. Baseline sessions consisted of reading the Iwata et al. (1982/1994) method; then, participants attempted to implement analog conditions. All employed a multi-component training program. Iwata et al. (2000) provided participants verbal and written descriptions of the conditions, followed by a video of a modeled functional analysis. Finally, during the reassessment of implementation, Iwata et al. provided trainees with feedback. Moore et al. (2002) also provided verbal and written descriptions, but they provided a live modeled functional analysis and the opportunity to rehearse the conditions prior to reassessment of implementation. Wallace et al. (2004) also provided descriptions of the conditions, video modeling, and rehearsal. If participants failed to meet the criterion, verbal feedback was provided.

Post-training, all participants were able to implement the functional analysis conditions with a high level of integrity, as shown by the criterion of 95% correct responding. Additionally, Moore et al. (2002) and Wallace et al. (2004) showed the increased procedural integrity generalized to functional analyses conducted with real clients.
TeachLivE offers the opportunity to train preservice and inservice teachers in a safe environment, both the FA procedures as well as data collection and analysis of data through a comprehensive intervention package. Therefore the purpose of this study is to replicate procedures from Kunnavatana, Bloom, Samaha, and Dayton (2011) and extend the literature using the innovative virtual learning environment platform TeachLivE to practice FA procedures with feedback for teachers.

Method

Participants and Setting

Participants were three preservice teachers who (a) had no previous experience with functional analysis implementation, (b) enrolled in a behavior analysis course, and (c) were willing to spend additional time beyond the existing course requirements to learn FA procedures.

Experimental Design and Procedure

A multiple-baseline, across-participants design was used to assess the effects of training on correct implementation of the attention, demand, and toy-play conditions.

Baseline. Simulated analyses sessions in which the participants acted as therapists conducting attention, demand, and toy-play sessions in the TeachLivE environment were five minutes in length, and an interactor assumed the role of an avatar who engaged in verbal tantrums. Prior to the sessions, actors were instructed to review scripts specifying the times at which target and non-target behaviors should occur, given the method section of the Iwata, Dorsey, Slifer, Bauman, and Richman (1982/1994) study.

Training. After baseline, the participants attended training with feedback/practice within the TeachLivE environment. In training, participants received a description and purpose of each functional analysis condition, demonstration of each condition within the virtual environment, and role playing within TeachLivE. Each participant alternated between playing the role of a data collector and the role of a therapist for all conditions. After role playing, presenters answered questions pertaining to the conditions with an After-Action-Review (AAR).

Results and Discussion

This study was designed to evaluate whether the promising technology of trial-based FA training in a virtual environment for teachers would yield results that could be used as the basis for effective function-based interventions for children who engaged in problem behaviors. In addition, we sought to provide sufficient detail such that readers could replicate these procedures with teachers. Although only three teachers participated, three functions were identified and treated. We replicated the procedural modifications by Kunnavatana et al. (2011) and demonstrated the effectiveness of the trial-based FA as an alternate method of FA within a virtual environment. Teacher training durations ranged from 45 to 90 minutes (M = 60 minutes). Thus,
the average total time invested in each of our trial-based FAs was just over 2.5 hours. Teachers conducted trials with an average of 96% procedural fidelity, suggesting that teachers were able to conduct trial-based FA with limited training in a virtual environment. However, our training was conducted individually, excepting the initial didactic presentation. This may not be the most economical approach for school districts that wish to adopt trial-based FA as part of their functional behavior-assessment protocols. Future researchers may wish to evaluate other approaches to training in which groups of teachers are trained simultaneously in virtual environments. Although we did not evaluate procedural fidelity using the teachers’ own data, the teachers’ procedural fidelity data agreed 100% with those of the independent observers. This suggests that teachers were able to detect their own errors, which may be helpful when they use trial-based FA after training.

References


Introduction

Teacher education institutions throughout the United States are preparing candidates for an increasingly diverse student population. Predominant among this changing population are English learners (ELs), projected to represent a quarter of K-12 students by 2025 (Goldenberg, 2008). Once the purview of English as a Second Language (ESL) specialists alone, English learners in the 21st century need all teachers to be equipped to support ELs’ academic subject achievement and language development.

The College of Education and Human Performance of the University of Central Florida has addressed the need to prepare teacher candidates of all subjects and grade levels to educate ELs in a comprehensive, systemic manner. Through implementing the One Plus Model of EL Infusion (Nutta, Mokhtari, & Strebel, 2012), the college has incorporated practice teaching cases of three ELs throughout various teacher preparation courses including: a dedicated course about teaching ELs (required of all teacher candidates), a foundational course, a general methods course, and subject-specific methods courses.

The EL cases studied throughout the infusion program were developed by the EL infusion coordinator after a year of weekly observation in an ESL classroom, interviews, and the collection of spoken and written language data of 40 students. Using the language data, and audio-recorded interviews with three students, the coordinator developed the cases of Edith, a newcomer from Mexico with a very minimal level of English proficiency, Edgar, an intermediate level English learner who left Puerto Rico eight months prior, and Tasir, an advanced English learner in the 7th grade who arrived in the US from Egypt in the 3rd grade. Faculty teaching courses that reference the three cases provide links to multimedia case descriptions, including language samples and audio-recorded interviews with students representing each level of English proficiency (for a detailed description of the cases, see Nutta, Strebel, Mokhtari, Mihai, & Crevecoeur-Bryant, 2014).

Over a three-year period, references to the cases expanded into more course activities and assignments, with candidates developing a concrete sense of what Edith, Edgar, and Tasir could comprehend and express in English. Most courses involved curriculum development, lesson planning and implementation, and creation of assessments embedded in the three cases, requiring adaptations for Edith, Edgar, and Tasir. Faculty and candidates found the cases relatable, memorable, and actionable, and candidates came to know the cases so well across multiple courses that they often brought them up even before their instructors did.

At the time these cases were taking hold in the teacher preparation curriculum, plans for the widespread application of the TeachLivE program using five native-speaker avatars (Ed, Sean, Kevin, CJ, and Maria) were in motion. The TeachLivE inventors approached the EL infusion
The development of the avatars drew upon the audio recordings and transcriptions of EL interviews and read-aloud sessions, informal and formal writing samples, video recordings of ELs participating in the ESL class and interacting socially with each other and with the teacher during breaks, and published research on second language acquisition (see Ellis, 2008) and learner language (see Gass, Behney, & Plonsky, 2013). Through these data, the avatar developers were able to accurately and authentically represent the comprehension ability and expressive capacity of each EL avatar, portraying the typical developmental patterns and common errors for each level of English proficiency.

The beginning student, Edith, can understand simple yes/no and either/or questions about the here and now (Teacher asks, “Is this Italy?” or “Is this Italy or Germany?” while pointing to Germany on a map); the intermediate student, Edgar, can understand more complex questions, including those that refer to the past and future (“How did President Hoover try to fight the Great Depression?”); and the advanced EL, Tasir, can understand most questions unless they are asked too quickly, with too much undefined academic language, or with infrequently used or idiomatic expressions. Expressive abilities and tendencies for Edith include frequently used one-word and short chunk phrases (“table” or “I like”) and reverting to Spanish when communication breaks down. Edgar can express himself in English with more complex phrases and sentences, but he makes frequent grammatical errors (“What the man doing in the picture?” “My brother, she 5 year old.”), and Tasir sounds very much like a native speaker, with no traces of foreign accent. She struggles with academic language, however, in her spoken and written expression, as well as in listening and reading comprehension.

The EL avatars have supercharged teaching cases by giving candidates the opportunity to interact with Edith, Edgar, and Tasir. Teacher candidates are taught to modify the language of their questions for Edith, Edgar, and Tasir using a questioning strategy called leveled questions (Nutta et al., 2014) and have the opportunity to practice their leveled questioning strategies with the EL avatars. No story, no video, and no microteaching experience can equal the immediate feedback received in real-time interaction with ELs at different levels of English proficiency. Candidates learn that even the best laid plans for level-appropriate questions can be challenged by unfamiliar words or sentence structures and that through interactional modifications (Ellis, 2008) in their communication with the avatars they can not only make instruction comprehensible but can also help build English language proficiency in their EL students.

**Steps and Facilitation**

TeachLivE’s English learner avatars have provided astonishing opportunities for pre-service and in-service teachers in the College of Education and Human Performance to practice their teaching skills. Currently, two of the core courses offered in multiple sections each semester utilize TeachLivE for the purpose of teacher preparation and practice. TSL4080 “Theories and Practices of Teaching ESOL Students in Schools” and EDG4410 “Teaching Strategies and Classroom Management” are both pioneers of the vast possibilities TeachLivE has to offer our teachers in training. The focus of this paper will be on the students who participated with TeachLivE in EDG4410.
The Courses: TSL 4080 and EDG 4410

In the course dedicated to the teaching of ELs, TSL4080, teacher candidates are introduced to the world of teaching English, become familiar with different levels of proficiency, and learn about various approaches they can pursue to meet their ELs’ needs. Prior to meeting and interacting with the avatars, the teacher candidates in this class are introduced to Edith, Edgar, and Tasir’s cases and learn about how each avatar represents a level of proficiency and background they will encounter as teachers in their mainstream classrooms. An hour-long leveled questions and verbal tasks workshop precedes the TeachLivE session to give the students more practice opportunities and knowledge on the types of questions they can ask ELs at different levels.

The general methods course, EDG 4410, is the second exposure to the EL teaching cases. The junior-level course is required of all elementary and secondary education majors. In the course, the teacher candidates explore instructional, organizational, and classroom management strategies to create effective learning environments. Included in the course requirements are a required 15-hour service-learning field experience and two microteaching sessions. Microteaching is a training technique whereby the participant reviews a recording of the lesson in order to conduct a "post-mortem." Participants find out what works, which aspects fall short, and what needs to be done to enhance their teaching technique. Invented in the mid-1960s at Stanford University by Dr. Dwight W. Allen, microteaching has been used successfully for several decades now, as a way to help teachers and teacher candidates acquire new skills (Peterson, 2000).

As the teacher candidates create their lesson plans for the microteaching sessions, they meet the three EL students, Edith, Edgar and Tasir, in the teaching cases via an online module. The lesson plans must contain ESOL accommodations for each EL with a detailed explanation of each accommodation and why the accommodation would help the EL at his or her level of language proficiency.

Procedures

After the microteaching session with their peers pretending to be K-12 students, the teacher candidates were given the opportunity to teach the lesson again in TeachLivE with the three EL avatars included in a class of five avatars. In preparation for the mixed-reality experience, the teacher candidates submitted their lesson plans electronically to TeachLivE. The teacher candidates were given an appointment for a five- to ten-minute time slot. The sessions took place in the TeachLivE lab with TESOL experts who observed and coached the teacher candidates. The coaches included two professors of ESOL education and one teacher educator who was trained by the ESOL professors in specific ESOL strategies. The three coaches rotated the observations using a rubric designed to indicate the ESOL strategies observed. After teaching the lesson, teacher candidates discussed their lessons with the TESOL experts and received feedback on their performance and recommendations on how they could improve their teaching skills. After the TeachLivE experience they were asked to write a reflection answering the following questions:

- How did you feel going into the lesson, knowing that at least one EL would be in the class?
- How did you feel when an EL student reacted to you? Why do you think you felt that way?
- Did you facilitate the interaction by using a planned EL strategy? If so, which strategy did you use? Was it successful? How do you know? Explain.
- Did you feel prepared to interact with the EL student? Explain.
Coaching Strategies

While observing the teacher candidates, the coaches focused on the candidates’ use of ESOL strategies using a rubric designed to indicate the ESOL strategies observed. The rubric consisted of a list of ESOL strategies called “show and tell” strategies. The coaches indicated on the rubric which show and tell strategies were observed, and wrote comments about each strategy. Show strategies are defined as strategies that provide nonverbal support for ESOL students in order to enhance their comprehension of verbal input. Tell strategies (verbal support) are those that involve the modification of the actual verbal or written input to enhance ESOL student comprehension (Nutta et al., 2014). Show and tell strategies can be used alone or in combination with one another. The show strategies observed most frequently included uses of visual materials such as photos, homemade pictures, posters, charts, graphs, and other realistic objects. The tell strategies observed most frequently included repetition of key words or phrases, rephrasing, simplification of the language of questions, and slower speech.

Coaching Takeaways

In conferencing with the candidates after their lesson presentations and reading the feedback from their written reflections, students shared their overall experiences and impressions of their self-efficacy in communicating with the ESOL students. In terms of their overall experience with TeachLivE, candidates overwhelming stated that the experience was very realistic and an excellent tool for teachers in training. Many candidates stated that every student in education should be required to use TeachLivE and others asked if they would have additional opportunities to do so. Regarding their self-efficacy with ESOL students, the participants remarked that they felt confident in their ability to interact with Edgar and Tasir, the intermediate and advanced ESOL students, but communicating with Edith, the ESOL student with the lowest English proficiency, was more challenging than they had imagined. Many candidates reported that they wished they had planned more strategies for Edith and for more individual interaction with all of the ESOL students. Finally, candidates reported that after their experience with TeachLivE, they learned the importance of using visuals, slowing their speech, and adjusting their questions to the appropriate level for their ESOL students.

Coaches agreed with candidate feedback in terms of their self-efficacy. Many candidates that were observed included visuals in their lessons, but far more were needed. Some students relied only on their verbal presentation for their entire lesson. Other candidates did not engage ELs in the lesson if they did not actively participate. Finally, some candidates struggled with the pacing and amount of speech. Common observations included candidates speaking too quickly, asking too many questions, or providing extremely wordy explanations for beginner or intermediate ELs.

References


Comparison of Feedback Methods for Pre-Service Teachers Delivering Mathematics Lessons in the TeachLivE Lab

Jacqueline Rodriguez
College of William and Mary

Abstract

Teacher preparation and development has recently come under scrutiny due to a wave of accountability measures put in place to focus on increased student outcomes. In order to prepare teachers for the accountability they will face in the classroom, teacher preparation programs are required to provide increased practice opportunities to pre-service teachers. School-based administrators and other educational leaders recognize that practical experience in the classroom, including internships, help prepare new teachers for classroom roles (Markow, Moessner, & Horowitz, 2006). Nevertheless, with accountability tied to high stakes testing, increased practice in public school classrooms is more difficult because teachers and administrators alike are concerned with pre-service teachers’ impact on student performance data. Therefore, strategic and careful considerations must be given to alternatives to the traditional brick and mortar classroom.

Background

Effective preparation programs not only provide practice opportunities for pre-service teachers, they also incorporate feedback to ensure pre-service teachers’ implementation of skills acquired (Simonson, Myers, & DeLuca, 2010). Feedback increases the likelihood of acquisition, implementation, and maintenance of skills across time and condition. Feedback methods have morphed with the changing generation of learners and newly developed technologies. Whereas face to face classroom observations were the norm in teaching practicums and yearly evaluations, new methods of feedback allow observers to participate in the process at a distance, all the while maintaining synchronicity of feedback. Bug-in-ear (BIE) technology allows observers to provide immediate, corrective feedback to teachers via voiceover internet protocols (VOIP). Many institutes of higher education and school-based administrators have considered using BIE technology for supervision and observation because it allows a teacher to immediately change a teaching technique instead of practicing an incorrect technique until delayed feedback is provided (Scheeler, McAfee, Ruhl, & Lee, 2006) while not causing any disruption to the classroom (Scheeler et al., 2009).

The use of praise to increase specific academic and social behaviors has a long history of significant effects (Partin, Robertson, Maggin, Oliver, & Wehby, 2010; Simonsen et al., 2010). Praise can be used to motivate and reinforce desirable behaviors (Partin et al., 2010). General and specific praise differ in their explicitness in explaining the reinforced behavior. In order for praise to be specific, the behavior must be explicitly defined for the student and praise must arise directly after the occurrence of the behavior (Simonsen et al., 2010). Similarly, general praise is verbal positive feedback, without the explicitness of the desired behavior (Simonsen et al., 2010).

Another effective strategy that has been found to increase positive academic behavior and decrease disruptive social behavior is the condition of increased opportunities to respond correctly to questions (Partin et al., 2010). The use of questioning in the classroom can increase engagement and discussion (Brophy & Good, 1986), it can be used to assess student learning (Stronge, 2007),
and it has been connected to facilitating discourse on content specific understanding in math (Piccolo, Harbaugh, Carter, Capraro, M., & Capraro, R., 2008) and science (Chin, 2006). Additionally, Sutherland, Adler, and Gunther (2003) associated increased opportunities to respond through questioning to increases in outcomes for students with disabilities and students from culturally and linguistically diverse backgrounds.

Research Questions

This study sought to determine if the after action review cycle (AARC) and corrective feedback delivered via bug-in-ear (BIE) demonstrated an impact in the virtual teaching environment. Therefore, the following three research questions were examined:

1. To what extent does the After Action Review Cycle (AARC) compared to Bug-In-Ear (BIE) feedback impact the frequency of academic questions asked by the participant?
2. To what extent does the After Action Review Cycle (AARC) compared to Bug-In-Ear (BIE) feedback impact the frequency of general praise disseminated by the participant?
3. To what extent does the After Action Review Cycle (AARC) compared to Bug-In-Ear (BIE) feedback impact the frequency of specific praise disseminated by the participant?

Method

Participants

Recruitment of participants took place in an undergraduate behavior management course in the College of Education of a large southeastern university. All students in the course were shown a presentation highlighting the study purpose, qualifications for participation, and perceived value of participation. Participation in the study was limited to students who were pre-service, undergraduate, exceptional education majors or minors completing their internship I practicum. Students were required to commit to a minimum of 13 teaching sessions in the virtual teaching lab on campus. Prior instruction or introduction to lab disqualified participation in the study. Three students were recruited as participants; however, the study sustained one case of attrition, which occurred before the intervention began. Of the two remaining participants, both were female, with an age range of 21-24. Participants were provided copies of the institutional review boards informed consent as well as schedules and procedures for the teaching sessions.

Setting and Materials

Interviews, training, and teaching sessions occurred in the same virtual teaching lab. The lab, TeachLivE, was located in a College of Education building on the university main campus.

Design

An alternating treatment design (ATD) was used to evaluate and compare the effects of the AARC and BIE feedback methods on teacher questioning and praise. Alternating treatments is defined as rapidly alternating two or more treatments at random (Horner, 2005).

Data collection

The participants in this study were naïve to the behaviors being measured: academic questioning and praise. Participants were given 10 minutes to teach a mathematics lesson focused
at the secondary level, specifically on middle school curriculum from the 6th-8th grade. Participants were sent the lesson plan via email 48 hours in advance to provide time for review. They were also allowed to make amendments to the procedures in the plan, but they were not permitted to change the lesson plan objectives. Both participants chose to use the lesson plans “as is” without making any noticeable changes to the teaching procedures.

Lesson plans were drawn from various educational agencies including: Discovery Education, The Annenberg Learner’s Insights into Algebra, and the Public Broadcasting System. Lessons focused on the 6th-8th grade National Council of Teachers of Mathematics standards. The study parameters required participants to teach a 10-minute direct instruction lesson in mathematics.

The AARC required the student to spend five minutes with the researcher to discuss the session. Based on critical components for feedback reviewed in Simonsen et al. (2010) and Colvin et al. (2009) the researcher would use specific examples throughout the session to discuss when appropriate questioning occurred, when there were missed opportunities for questioning based on student feedback and interactions, and then the researcher and the participant would review the lesson plan to identify alternate areas where questioning was appropriate. An identical procedure was used to discuss general and specific praise.

The BIE feedback method requires the student to adapt their teaching in real time. The teacher must multi-task their interaction with the curriculum, with the students, and with the researcher-prompted feedback. Feedback during this intervention was delivered synchronously, supporting only rapid, short statements to illicit student responses, for example, “Ask Marcus to define perimeter,” or “Remember to praise Maria for effort”. Participants receiving BIE feedback did not participate in a five-minute review after the session.

Inter-observer agreement

An experienced teacher in the surrounding school system acted as inter-rater observer. The teacher has over 10 years of combined experience teaching secondary and elementary students. Agreement across behaviors was calculated by dividing the number of agreements minus the number of disagreements, by the total number of agreements and multiplying by 100%. In total, inter-observer agreement was collected on (61%) intervention sessions, with a mean agreement of 97.5% across participants on the occurrence of questions and praise.

Inter-observer agreement was also collected for treatment integrity. Agreement data was collected on 61% of AARC sessions using a checklist. Treatment integrity was 95.8% across all sessions.

Results

Three research questions were examined throughout this study: To what extent does the After Action Review Cycle (AARC) compared to Bug-In-Ear (BIE) feedback impact the frequency of academic questions asked by the participant? To what extent does the After Action Review Cycle (AARC) compared to Bug-In-Ear (BIE) feedback impact the frequency of general praise disseminated by the participant? To what extent does the After Action Review Cycle (AARC) compared to Bug-In-Ear (BIE) feedback impact the frequency of specific praise disseminated by the participant?
“Percentage of non-overlapping data (PND) is the most critical statistic to report when comparing conditions with the ATD” (Gast, 2010, p. 352). Whereas variability in data paths is very important in other single subject designs, ATD is most interested in the changes in one data point that results in the superiority of one of the conditions. Scruggs and Mastropieri (1998, 2001) suggest using a range of PND scores to define effective treatments. Scores above 90 are considered very effective, while scores between 70 and 90 are considered effective; scores between 50 and 70 are considered questionable, while scores below 50 are considered ineffective (Scruggs & Mastropieri, 2001).

Results of this study indicated that across participants, BIE was the most effective treatment to increase general praise in the classroom ($M_{PND} = 100\%$). Effectiveness (Scruggs & Mastropieri, 2001) of BIE treatment to increase specific praise across participants was questionable ($M_{PND} = 62.5\%$), as was BIE treatment to increase academic questioning ($M_{PND} = 50\%$). In examination of each participant, BIE treatment was highly effective in increasing general praise ($M_{PND} = 100\%$) and effective in increasing specific praise ($M_{PND} = 75\%$) for Participant 2.

Discussion

Limitations

Replication of this study would require researchers to consider several limitations to the current study. This study did not include a baseline for all participants or a maintenance period to determine worth of the intervention. Further proceduralizing of the feedback provided to the participant during both interventions will control for variability of the personnel providing the feedback. Coaching to a criterion and collecting data on whether the participant met the criteria may provide a more comprehensive picture of the effectiveness of the intervention.

Carryover effect is a confounding factor in multitreatment designs. In future studies, the study schedule should be spaced out to provide consistent, ample time in between sessions. Although multiple treatment sessions occurred on the same day, this study did not find any carryover effect. Finally, results of this study were marred by teaching behaviors not controlled during participant recruitment. For example, managing classroom behavior was assumed as both participants were synchronously enrolled in a behavior management course during the course of the study. However, managing off-task and disorderly behavior disrupted the flow of the lesson, causing periods ranging from 2-5 minutes that the participant spent remediating a student, and asked less than 5 content related questions with little to no opportunity for student praise.

Overall Impact

Although a clear separation of treatment effects was demonstrated in assessing general praise, there is not enough evidence to suggest that either treatment was more effective in increasing specific praise or questioning. Furthermore, the discreet values indicate a nominal change in the frequency of general praise. There is not enough data to suggest a more effective feedback method when evaluating AARC versus BIE for pre-service teachers.
References


A Comparison of Virtual Learning Environments on the Social Responses for Children with Autism

Eleazar Vasquez III, Ph.D. Gregory Welch, Ph.D. Athira Alias
Carrie Straub, Ph.D. Charlie Hughes Ph.D. Melissa Paz
Arjun Nagendran, Ph.D. Morgan Russell Kelly Schaffer, Ph.D.
Matthew Todd Marino, Ph.D. Aaron Koch Jessica Silva
Lauren Delisio

University of Central Florida

Introduction

Individuals with autism spectrum disorders (ASD) are characterized by a marked difference in ability to communicate and form relationships. As of 2014, the Centers for Disease Control and Prevention estimated 1 in 68 children had been identified with ASD. Differences in communication and socialization abilities, such as an awareness of other people’s feelings, along with engagement in repetitive behaviors, often appear in early childhood and continue throughout life (American Psychiatric Association [APA], 2013).

New technologies for communication and learning have been used to improve social communication for children with ASD. Interactive, virtual environments (VEs) are three-dimensional, real-time, computer-based representations of reality that simulate the real or imagined world. A well-known example is Second Life, yet there are a variety of VE settings. Equipment for VEs varies in expense and level of fidelity to reality, employing a range of displays from head mounted displays to desktop computer VEs. Sensory information is generally provided with audio or video feed; however, movement and touch can also be incorporated. Users generally interact with the environment via an avatar, or digital representation of the user. Avatars in VEs can provide a first person or third person perspective, and VEs have varying levels of a sense of immersion, or “presence”. This sense of presence allows the user to perceive the VE as an authentic environment in which action can be taken, much like in the real world. However, VEs provide benefits which are unlike the real world: users can engage in activities without experiencing long-term consequences of their actions, allowing them to engage safely in practice and learn from mistakes in a virtual environment. Vera, Campos, Herrera, and Romero (2007) contended that VEs are uniquely suited for learning because: a) they remove competing stimuli which can distract or overwhelm the user, b) learning time can be segmented with short breaks that allow the instructor to clarify factors, and c) they involve a perceived element of play which engages users. Thus, VEs provide a safe and potentially engaging avenue for individuals in virtual environments to learn skills that can be approximated in a computer-based simulation.

While new venues for delivering social skills instruction are emerging, it is not clear whether or not students with ASD will engage in social interactions in the aforementioned VEs. Carter, Williams, Hodgins, and Lehman (2014) observed twelve 4- to 8-year-old children with ASD as they interacted with a human therapist, a VA controlled by an interactor, a human actor using vocal characterizations of the VA, and pre-recorded cartoon characters which elicited responses from the viewing audience after delivering a prompt. Carter et al. found a significant main effect of condition on interactions, with the therapist condition resulting in the highest proportion of appropriate responses [.42(.13)], followed by the actor [.24(.10)], the VA [.23(.18)],
and the cartoon characters [.07(.09)]. Their findings suggested that children with ASD were more responsive to a human than a VA. However, it is notable that the therapist in the study was a doctoral level speech-language pathologist with over 30 years of experience with young children with developmental language issues, while the VA interactor had no specialized training for children with ASD. Also, in all conditions but the VA, the interactions were one-to-one, and the VA interactions took place with participants as members of an audience with other children present. Carter et al. suggested future research should incorporate one-to-one interactions across all conditions.

The current study compares social interactions of children with ASD in multiple environments, real and virtual, across three modalities: a) human, face-to-face interactions (human condition), b) virtual avatar (VA), and c) physical-virtual avatar (PVA), specifically addressing the following questions:

1. To what extent does student-initiated social responding compare across three modalities (human, VA, and PVA) for adolescents with autism?
2. To what extent does student-reciprocal social responding compare across the three modalities?
3. How do the mean scores of duration of total social responding compare across the three modalities?

### Method

#### Participants and Settings

Participants, three adolescent males between 15 and 16 years of age, were included based on a previous diagnosis of ASD by a licensed psychologist and were selected from the following criteria as reported by parents: a) knows his own name, b) responds to ‘No’ or ‘Stop’, c) knows 10 or more words, and d) can use sentences with four or more words. Participants did not use assistive technology (i.e., voice output devices or pictorial communication) as primary means of verbal communication. Christian and Nicholas were enrolled in ninth grade, Brian was enrolled in tenth grade, and each received special education services in the public school system.

The study was conducted in a university simulation lab with three private rooms connected by a hallway and waiting area. The human condition room contained one Caucasian adolescent male seated at a conference table with an empty chair. The table was in front of a white board and no other significant items were located on the walls of the room. The VA condition room contained a high-definition display monitor mounted approximately 36 inches from the floor and speakers. The display and speakers were used as the delivery platform for the VA social interaction. Partitions were set up around the display monitor, leaving approximately 10 feet by 10 feet of space, blocking any extraneous stimuli from vision. The PVA condition room contained a life-sized animatronic entity used as the delivery platform for the PVA technology social interaction. A video recorder was placed in each room to record data for later analysis.

#### Results and Discussion

Social interactions of three adolescents with ASD were analyzed across three conditions to detect differences: (a) human, (b) VA, and (c) PVA. Across all three participants, more social interactions were initiated with the avatars versus the human, with two individuals initiating more
and the cartoon characters [.07(.09)]. Their findings suggested that children with ASD were more responsive to a human than a VA. However, it is notable that the therapist in the study was a doctoral level speech-language pathologist with over 30 years of experience with young children with developmental language issues, while the VA interactor had no specialized training for children with ASD. Also, in all conditions but the VA, the interactions were one-to-one, and the VA interactions took place with participants as members of an audience with other children present. Carter et al. suggested future research should incorporate one-to-one interactions across all conditions.

The current study compares social interactions of children with ASD in multiple environments, real and virtual, across three modalities: a) human, face-to-face interactions (human condition), b) virtual avatar (VA), and c) physical-virtual avatar (PVA), specifically addressing the following questions:

1. To what extent does student-initiated social responding compare across three modalities (human, VA, and PVA) for adolescents with autism?
2. To what extent does student-reciprocal social responding compare across the three modalities?
3. How do the mean scores of duration of total social responding compare across the three modalities?

Method

Participants and Settings

Participants, three adolescent males between 15 and 16 years of age, were included based on a previous diagnosis of ASD by a licensed psychologist and were selected from the following criteria as reported by parents: a) knows his own name, b) responds to ‘No’ or ‘Stop’, c) knows 10 or more words, and d) can use sentences with four or more words. Participants did not use assistive technology (i.e., voice output devices or pictorial communication) as primary means of verbal communication. Christian and Nicholas were enrolled in ninth grade, Brian was enrolled in tenth grade, and each received special education services in the public school system.

The study was conducted in a university simulation lab with three private rooms connected by a hallway and waiting area. The human condition room contained one Caucasian adolescent male seated at a conference table with an empty chair. The table was in front of a white board and no other significant items were located on the walls of the room. The VA condition room contained a high-definition display monitor mounted approximately 36 inches from the floor and speakers. The display and speakers were used as the delivery platform for the VA social interaction. Partitions were set up around the display monitor, leaving approximately 10 feet by 10 feet of space, blocking any extraneous stimuli from vision. The PVA condition room contained a life-sized animatronic entity used as the delivery platform for the PVA technology social interaction. A video recorder was placed in each room to record data for later analysis.

Results and Discussion

Social interactions of three adolescents with ASD were analyzed across three conditions to detect differences: (a) human, (b) VA, and (c) PVA. Across all three participants, more social interactions were initiated with the avatars versus the human, with two individuals initiating more
interaction with the VA and the other with the PVA. In terms of responses, two of the three participants responded more frequently to the avatars than the human, while a clear pattern did not emerge with third participant. Finally, all three participants on average had longer conversations with the avatars, with the VA lasting the longest, followed by the PVA, then human. Although this study includes a small number of participants, the alternating treatments design demonstrated a replication of results within and across subjects, lending to internal validity of findings. Further, the design of the study and the protocols enacted in the one-to-one interactions allowed for a more controlled setting than in previous research with VAs.

**Figure 1.** Distribution of Initiated Social Interaction Across Environments.

**Figure 2.** Distribution of Reciprocal Social Interaction Across Environments.
Findings in the present study support Baron-Cohen’s (2009) systemizing theory that individuals with ASD prefer interacting with digital characters, which have faces that are less complex than human faces. It may be that faces of digital characters, avatars, are more easily comprehended by individuals with ASD. However, results do not support the findings of Carter et al. (2014), who found increased social interactions with a human therapist when compared to a VA. Further, this is the first study to the author’s knowledge suggesting students with ASD will initiate social interactions with surrogates. There may be many possible reasons for the difference in results: the children were of a younger age range in Carter et al.’s study; children interacted with the therapist in a one-to-one interaction, yet with the VA in a one-to-many interaction; and the human was a trained professional with over 30 years of experience. Carter et al. suggested that a trained professional interacting as a VA may have the most positive results, and our research supported this hypothesis also. Future research should involve randomized controlled trials of matched participants who communicate in different conditions, including with a VA controlled by a professional with experience with students with ASD, a VA controlled by a professional actor, and a human. The findings of this study suggest that adolescents with ASD may initiate and respond more to interactive, digital characters, and that their conversations may last longer than with humans. Results of this study can be used to shape social skills curricula in VEs with students with ASD.

References

American Psychiatric Association. (2013). The Diagnostic and Statistical Manual of Mental Disorders: DSM 5. bookpointUS.


Job Interview Coaching in a Virtual Environment for Individuals with Intellectual Disabilities

Zachary M. Walker                   Eleazar Vasquez III                  Wilfred Wienke
National Institute of Education     University of Central Florida         University of Central Florida

Introduction

Preparing students and families to cope with the challenges of transitioning into postsecondary settings is a complex process for any student, especially for students with disabilities. This difficulty is further substantiated by poor employment outcomes for students with disabilities in the United States. Fifty-seven percent of youth with disabilities are likely to work once they complete secondary schooling, as compared to 66% of the general population (Newman, Wagner, Cameto, & Knokey, 2009). The Bureau of Labor Statistics (BLS) (2013) reports that the highest percentages of unemployed persons come from two groups: high school dropouts and people with disabilities. In April 2013, the percentage of unemployed individuals with disabilities was 12.9%, while the percentage of unemployed individuals without disabilities was 6.9% (BLS, 2013). Even with a high school diploma, an individual with a disability is far less likely to obtain a job than someone without a disability. For example, only 58% of disabled individuals are employed full-time up to four years out of high school, and the majority of those individuals report having to work two to three part-time jobs to meet full-time hours (Newman et al., 2009).

These data are alarming, with implications beyond the financial well-being of the disabled and their families. The effects of unemployment are much greater than lack of income, and can have a significant negative effect on happiness and life-satisfaction (Kassenboehmer & Hasisken-DeNew, 2009). Employment has a great impact on quality of life; however, everyone needs prerequisite skills for successful employment. While certainly not the only reason for the dismal employment outcomes for those with disabilities, one such skill is interview performance.

Purpose of the Study

The purpose of this study was to identify if functional relations exist between a treatment package of interview practice in a virtual learning environment (TLE TeachLivE™) and interview performance coaching for young adults with ID. The intervention was delivered as a package and no attempt was made to analyze the contribution of the separate components. Specifically, we asked the following three questions:

1) To what extent will the combination of interview practice in the TeachLivE lab and coaching increase job interview performance for 18- to 22-year old participants with intellectual disability as measured by an interview rubric?
2) Will social skills in during job interviews, as demonstrated following the combination of interview practice in the TeachLivE lab and coaching, transfer to a live simulated job interview for young adults ages 18-22 with intellectual disability?

3) How socially important do participants, parents/primary caregivers, and employee experts rate the goals, procedures, and outcomes of this research study as measured by a survey?

While virtual learning environments have been associated with gains in academic skills for certain groups, the ability to increase social skills in an interview setting would be a novel and potentially powerful use of virtual environments.

This study included five 18- to 22-year-old participants. In order to participate in the research study, participants needed to be enrolled in a large public school transition program for 18- to 22-year-olds who have not yet received a high school diploma. The transition program entails classroom instruction on a university campus and work internships both on-campus and in the community. Participants possessed an IQ within the range of 55 to 65, and could not have had more than two years of paid work experience as documented on a completed employment history questionnaire as part of the transition program. In an attempt to ensure regular participation in the study, selection criteria for participants included consistent and regular school attendance (less than eight absences for the prior semester).

The first location for this study was the TeachLivE virtual classroom laboratory, for participants to practice interview skills in a real-time, mixed-reality setting. Participants took part in both baseline and intervention treatments in this setting. During baseline and treatment interviews, the participant sat facing the television. The space was a windowless room with three beige colored walls and one green wall. A large projection screen was located slightly left of the center of the room, and was roughly 12 feet from the entryway. A 70-inch high-definition flat screen television was suspended approximately three feet from the floor in front of the screen. A screened space adjoined the projection screen on the left-hand side and provided a divider for an on-site TeachLivE technician to assist in program operations. A logistics webcam mounted on the top of the projection screen allowed the interactor to view the participant during sessions, and speakers behind the screen enabled the interactor to hear the participant. Real-time communication between the interactor and the participants occurred via Skype. The professional interactor was in control of the behavior of the avatar from a remote setting. The interactor was a trained improvisational actor with three years of experience working in the TeachLivE lab. The second setting was a small classroom (15 feet by 21 feet, containing a round table and chairs) adjacent to the TeachLivE lab, where coaching sessions were conducted following treatment interview sessions.
Results and Discussion

The outcomes of this study suggest the combination of interview practice in the simulated virtual environment and coaching sessions resulted in dramatic increases in interview performance in live settings for adults with ID. The use of mixed-reality environments and coaching to provide instruction for individuals with disabilities provides many innovative possibilities for further research. Mixed-reality environments can act as a medium for instruction and practicing behaviors while the coaching functions as the instruction itself. The particular type of instruction employed by a teacher (e.g., direct instruction, constructivist method, etc.) could be used in any setting. Mixed-reality provides the unique opportunity for individuals to practice these skills in a realistic setting that does not result in harm to the participant or the “practice partner”, since the avatars are not real (Dieker et al., 2008).

In regards to this study, it will be interesting to investigate if interview practice in a mixed-reality environment is the most significant factor in altering interview performance, or if the utilization of coaching increases or decreases participant performance. The effect of each variable could be analyzed by comparing interview performance after practice interviews with no coaching to interview performance after coaching sessions with no practice. The combination of variables was successful in this study, but further research would be needed to identify to what degree each component improved performance.

Further research should also study the effects of specific characteristics (e.g., gender, age, and dress of the avatar or the participants) on treatment outcomes. For this study, we used one middle-aged female avatar, dressed very formally, practicing a set collection of questions. Further development of avatars will be helpful in future studies, as results could vary with differentiation of an avatar’s age, gender, style of dress, and manner of questioning. The coach in this study was a retired female teacher; perhaps results would have varied with a coach of a different age, gender, or personality. Additionally, the study included only one male participant. Results may differ with more males introduced to the study. Changing the variables and determining the most effective arrangements, while a challenge, may be useful.

Future research may also test the reliability and compare the validity of other evidence based models of instruction (e.g. direct instruction, video modeling, etc.). For example, would results improve if we added a video modeling component to instruction? Would results occur sooner or generalize differently if a different type of instruction is used? It may also be useful to study the combination of video modeling and coaching before practicing in the mixed-reality environments.
References


Examining Potential Teacher Bias of Hispanic Males with Emotional Disturbances in Virtual Settings

Angel L. Lopez, Jr.
Ph.D. Dissertation
University of Central Florida

Abstract

The holistic evaluation of student and teacher interactions is essential to reflective teaching practices; consequently, the reflective practice of teaching needs to incorporate all facets of the teacher, known and unknown. This study looks at the potential influence of hidden biases towards adolescent Hispanic males and students with Emotional Behavior Disorders by observing preservice teacher (PT) interactions with students within a simulated classroom environment. Factorial MANOVAs and discriminant analyses revealed statistically significant interactions and relationships between participant level of bias and the identified student-avatars. These exchanges were more prevalent with one student-avatar by both experimental and control PTs, indicating that student characteristics and their differences are important factors in addressing issues related to bias.
Individualized Clinical Coaching in an Interactive Virtual Reality to Increase Fidelity of Implementation of Discrete Trial Teaching

Krista Vince Garland
Ph.D. Dissertation
University of Central Florida

Abstract

Discrete-trials teaching (DTT) is an evidence-based practice in educational programs for children with autism spectrum disorders (ASD). Despite a strong demand for effective teacher implementation of DTT, published research on such studies is scarce. A multiple-baseline, across-participants design was utilized to determine a possible functional relationship between individualized clinical coaching within the TeachLivE™ virtual reality classroom and teachers’ fidelity of implementation of discrete trial teaching when working with a student with ASD in an actual classroom setting. Following intervention, participants’ DTT accuracy improved on average from 14% to 80%, a 66% increase over a range of four to seven 15-minute sessions. Generalization probes indicate that participants increased their fidelity in delivering DTT in the classroom from 10% to 90% on average. These results indicate that such training has considerable potential to teach educators to implement DTT best practices in their classrooms with sustained fidelity. Implications and suggestions are discussed for future teacher preparation research utilizing individualized clinical coaching within a virtual classroom.
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 TeachLivE™ 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 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 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 19 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 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 Question 1:** Are there differences in performance over four 10-minute sessions of TeachLivE in a classroom simulator?

**Research Question 2:** Are there differences in performance over four 10-minute sessions based on whether or not teachers received 40 minutes of online PD?

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>TLE &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>
<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>
</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></td>
<td>$n$ (%)</td>
<td>$n$ (%)</td>
<td>$n$ (%)</td>
<td>$n$ (%)</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>

Valid Participants = 135     Attrition = 22

**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>10-minute session</td>
</tr>
<tr>
<td></td>
<td>45 –95 minutes</td>
<td>(4 observations)</td>
</tr>
<tr>
<td></td>
<td>(2 observations)</td>
<td></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</td>
<td>Frequency per</td>
</tr>
<tr>
<td></td>
<td>1 minute interval</td>
<td>10-minute session</td>
</tr>
<tr>
<td></td>
<td>every 5 minutes</td>
<td></td>
</tr>
</tbody>
</table>

Sub-constructs from 2011 8 sub-constructs

Danielson Framework for Teaching

Type of Data Collected - 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.

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 teacher schedules 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. 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, 11 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></th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1: Teachers oriented and schools randomly assigned</td>
<td>34</td>
<td>41</td>
<td>40</td>
<td>42</td>
<td>157</td>
</tr>
<tr>
<td>Step 2: Changes in groups per teacher request</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From G1 to G3</td>
<td>-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From G2 to G1</td>
<td></td>
<td>3</td>
<td></td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td>From G3 to G1</td>
<td>2</td>
<td></td>
<td>-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>From G4 to G1</td>
<td></td>
<td>1</td>
<td></td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>From G4 to G3</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
From G4 to G2

<table>
<thead>
<tr>
<th>Step 3: Teachers lost to attrition</th>
<th>-2</th>
<th>-4</th>
<th>-7</th>
<th>-9</th>
<th>-22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 4: Final number of teachers per group</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>30</td>
<td>135</td>
</tr>
</tbody>
</table>

**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 content. During TeachLivE sessions, the facilitator followed a detailed procedural checklist to operate 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>DE Pre</td>
<td>r</td>
<td>1</td>
<td>.130</td>
<td>.292**</td>
<td>.371**</td>
<td>.220'</td>
</tr>
<tr>
<td>DE Post</td>
<td>r</td>
<td>.130</td>
<td>1</td>
<td>-.028</td>
<td>.259**</td>
<td>.353**</td>
</tr>
<tr>
<td>SF Pre</td>
<td>r</td>
<td>.292**</td>
<td>-.028</td>
<td>1</td>
<td>.183'</td>
<td>.199'</td>
</tr>
<tr>
<td>SF Post</td>
<td>r</td>
<td>.208</td>
<td>.259**</td>
<td>.076</td>
<td>1</td>
<td>.118</td>
</tr>
<tr>
<td>TPOT Sum Pre</td>
<td>r</td>
<td>.371**</td>
<td>.185*</td>
<td>.183*</td>
<td>1</td>
<td>.633**</td>
</tr>
<tr>
<td>TPOT Sum Post</td>
<td>r</td>
<td>.220'</td>
<td>.353**</td>
<td>.171</td>
<td>.199'</td>
<td>.633**</td>
</tr>
</tbody>
</table>

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

**. Correlation is significant at the 0.01 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 and SF. 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 found 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>
<thead>
<tr>
<th>PD Factor</th>
<th>TeachLivE Sessions</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Session 1 M (SD)</td>
<td>Session 2 M (SD)</td>
<td>Session 3 M (SD)</td>
<td>Session 4 M (SD)</td>
<td></td>
</tr>
<tr>
<td>No Online PD</td>
<td>34</td>
<td>5.1 (4.2)</td>
<td>7.6 (4.9)</td>
<td>8.4 (5.3)</td>
<td>9.9 (4.8)</td>
</tr>
<tr>
<td>Online PD</td>
<td>25</td>
<td>6.5 (4.2)</td>
<td>7.9 (4.5)</td>
<td>7.9 (5.9)</td>
<td>9.5 (8.1)</td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>5.7 (4.2)</td>
<td>7.8 (4.7)</td>
<td>8.2 (5.5)</td>
<td>9.8 (6.3)</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>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Session 1</td>
<td>Session 2</td>
<td>Session 3</td>
<td>Session 4</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>
<td>8.6 (4.6)</td>
</tr>
<tr>
<td>Online PD</td>
<td>24</td>
<td>6.9 (4.5)</td>
<td>6.7 (3.6)</td>
<td>6.3 (3.9)</td>
<td>7.9 (6.7)</td>
</tr>
<tr>
<td>Total</td>
<td>58</td>
<td>6.5 (4.8)</td>
<td>7.8 (5.1)</td>
<td>7.7 (4.6)</td>
<td>8.3 (5.5)</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
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. No outliers were identified by inspection of a boxplot. There was homogeneity of variances for percentage of DE asked 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 \( (F(1,130) = .002, p = .968, \eta^2_p = .000) \), 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 \( (F(1,130) = .304, p = .582, \eta^2_p = .002) \) or for time \( (F(1,130) = 1.580, p = .211, \eta^2_p = .012) \).

**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. Only one outlier existed in the data, which was in G2 (online PD factor with no TLE factor) as assessed by inspection of a boxplot, and the outlier was not modified or removed from the data set. 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 \( (F(1,130) = .168, p = .682, \eta^2_p = .001) \). There was a statistically significant two-way interaction between time and TeachLivE \( (F(1,130) = 3.479, p = .064, \eta^2_p = .026) \) and time and online PD \( (F(1,130) = 5.735, p = .018, \eta^2_p = .042) \). The interaction between TLE and online was not statistically significant \( (F(1,130) = .015, p = .902, \eta^2_p = .000) \). Statistical significance of a simple main effect was accepted at a Bonferroni-
adjusted alpha level of .05. There was a statistically significant simple main effect of online PD at pre-intervention ($F(1,130) = 4.854, p = .029, \eta^2_p = .036$), but not at post-intervention ($F(1,130) = 1.204, p = .902, \eta^2_p = .036$). Conversely, there was not a statistically significant simple main effect of TeachLivE at pre-intervention ($F(1,130) = 1.274, p = .261, \eta^2_p = .010$), but there was a post-intervention ($F(1,130) = 9.827, p = .002, \eta^2_p = .070$). 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. Mean percentage DE was higher at pre-intervention for those assigned to the online PD groups than those who were not, with a mean difference of 5.7% (90% CI, 0.014 to 0.100), $p = .029$. However, 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), $p = .002$. Teachers who received the online PD decreased their questions by 3%, whereas those who did not increased questions by 7%. Conversely, TeachLivE teachers increased DE questions by 6%, whereas teachers who did not get TeachLivE decreased them by 2%. 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>TeachLivE</td>
<td>No TeachLivE</td>
</tr>
<tr>
<td>Pre</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</td>
<td>-2</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. Three outliers existed in the data as assessed by inspection of a boxplot. All three outliers were at pre-intervention, each with a value of 100% SF given in a class (outliers were in G2, G3, and G4, one per group; i.e., all groups except for the group with no between-subjects factors). None of the outliers were modified or removed from the data set. 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 ($F(1,130) = 3.486, p = .064, \eta^2_p = 0.26$). 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(1,130) = .527, p = .469, η2p = .004). Statistical significance of a simple main effect was accepted at a Bonferroni-adjusted alpha level of .050. There was a statistically significant simple main effect at pre-intervention for groups with TeachLivE as a factor, G3 and G4 (F(1,131) = 4.411, p = .038, η2p = .033), but not for those without TeachLivE, G1 and G2 (F(1,131) = .321, p = .572, η2p = .002). 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. 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, those receiving TeachLivE without online PD, 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>
<thead>
<tr>
<th>Table 9. Changes in Mean Percentages of SF over Time.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment Groups</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Comparison</td>
</tr>
<tr>
<td>Online PD</td>
</tr>
<tr>
<td>TeachLivE</td>
</tr>
<tr>
<td>TeachLivE &amp; Online PD</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. In G1,
three outliers existed at pre-intervention. In G2, two outliers were in pre-intervention and one was at post. In G3, two outliers were at pre-intervention and one was at post. In G4, two outliers were at pre-intervention and one was at post. No outliers were modified or removed from the data set. 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, η²_p = .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, η²_p = .009), nor post-intervention (F(1,121) = .008, p = .928, η²_p = .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>Time</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></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison</td>
<td>32</td>
<td>22.06</td>
<td>(3.75)</td>
<td>22.00</td>
<td>-.06</td>
</tr>
<tr>
<td>Online PD</td>
<td>32</td>
<td>21.33</td>
<td>(5.35)</td>
<td>21.83(4.81)</td>
<td>+.50</td>
</tr>
<tr>
<td>TeachLivE</td>
<td>32</td>
<td>21.63</td>
<td>(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</td>
<td>(3.88)</td>
<td>22.41(4.49)</td>
<td>-.78</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. TPOT Sum 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. No outliers existed in the data. 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 ($F(1, 194) = .087$, $p = .768$, $\eta^2_p = 000$). 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>
<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>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 ($F(3, 130) = 3.479, p = \cdot064, \eta^2_p = .026$) 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, 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

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

**Teacher Code:** __________  **Date:** __________  **Observer's Initials:** __________

<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.

**Establishing a culture for learning:**

- T. communicates expectations for lesson
- T. encourages students "do your best".

**Managing student behavior:**

- lack of commitment to learning
- little commitment to learning
- high expectations by teacher
- shared belief in importance

**Managing classroom procedures:**

- no established standards of conduct
- inconsistent standards of conduct
- teacher established standards
- students self-monitor with standards

<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 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

<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</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wait Time</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>1 min</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.

### 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

<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</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wait Time</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Field Log</strong></td>
<td><strong>2 min</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Write at least one note about practice per area then check appropriate box. Provide comments to justify rating.*

#### 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

<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</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wait Time</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Field Log</strong></td>
<td><strong>2 min</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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

<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</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wait Time</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Field Log</strong></td>
<td><strong>2 min</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Write at least one note about practice per area then check appropriate box. Provide comments to justify rating.*

#### 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)
### 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</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wait Time</td>
<td>1 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>1 min</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.

| 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 |
| 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 |
| 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) |
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:

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

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 have added five to both sides, like she did.

Ben: Ben did not do the opposite of substituting, he should have added 5 to both sides.

Show how you would work it out:

\[
\begin{align*}
5 - x &= 6 \\
5 + 5 &= 6 + 5 \\
x &= 11
\end{align*}
\]
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: 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 output could never be true.

Show how you would work it out:

You take 11 in place of \( x \) and substitute it by 5. If you would get 6 as the output then \( 5 - 11 = 6 \) true.
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 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
s - s
1 - s = 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:

$$5 - x = 6$$

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

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:

...