Iowa Initiative for Artificial Intelligence
Final Report

Project title:	**Immersive Virtual Mathematics Education for Teachers (Project VIME)**
Principal Investigator:	Seth King & Anne Estapa
Prepared by (IIAI):	Tyler Bell
Other investigators:	
Date:	2020-2022
Were specific aims fulfilled:	Y
Readiness for extramural proposal?	Y
If yes … Planned submission date	2021 (NSF); 2022 (IES)
Funding agency	See above
Grant mechanism	Research on Emerging Technologies for Teaching and Learning (NSF); Special Education Research Competition (IES)
If no … Why not? What went wrong?	

<u>Brief summary of accomplished results:</u>

The goal of our project was to create an automated 'smart' virtual reality (VR) simulation capable of administering assessment and providing feedback as part of an evidence-based personnel training strategy relative to the performance of a verbal mathematical questioning strategy. Accomplishing project aims involved the creation of an interactive VR classroom environment and the use of several artificial intelligence (AI) services, including speech-to-text, text-to-speech, and natural language classification. Additionally, we developed a web application to allow for the creation of lesson content and review of trainee performance. Our series of experiments demonstrated the effectiveness of smart VR for complex instructional activities and illustrated the feasibility of an iterative approach to development incorporating single-case and group experimental designs. Over the course of the funded project period, we successfully designed and evaluated the effectiveness of the proposed simulation via two studies. Future work will aim to incorporate additional features into the simulation (e.g., motion capture animation, assessment of trainee motion) while addressing a broader range of practices in education and other disciplines.

<u>Research report:</u>

Aims: Original aims of the project include the following:

<u>Aim 1</u>: Develop a pilot version of an immersive, AI-augmented VR simulation that independently generates a simplified instructional scenario focused on multiplication and assesses implementation of evidence-based prompting in a mathematics context.

<u>Aim 2</u>: Compare the efficacy of the simulation, relative to those who received typical professional development.

<u>Aim 3</u>: Demonstrate cascading effects of the training on mathematical performance of students with ASD. *(Note: We ultimately did not pursue work involving teachers or students with autism, as the COVID-19 pandemic made accessing these populations difficult. This decision was made unanimously by the project team.)*

Data:

Project members generated a set of training data that contains example educator responses to a student learning how to perform simple multiplication. In this exercise, educator responses may fall into a variety of categories. The training data consists of several (~10) responses for each response category. This dataset allowed a natural language classifier to be trained such that an educator's response to a student could be classified, allowing the developed system to determine if their response matched best practice.

AI/ML Approach:

Experimental methods, validation approach:

The initial aim of this research was to create a customizable 'smart' VR simulation for conversation-based training. This was achieved by first creating a web application that allows instructors or subject matter experts to generate instructional content—in the form a flowchart—for the simulated trainings. A VR-based simulation was then developed to retrieve the training content and generate automated "lessons" for trainees to practice and be evaluated within. The VR simulation was developed in the Unity game engine and was deployed on the Oculus Quest 2 VR headset. To analyze user interactions and give feedback to the trainee, AI services were used, including those for speech-to-text, natural language classification, and text-to-speech. The simulation was able to send data to the AI services (IBM Watson, Google Natural Language), analyze the user's verbal response to see if it corresponded to best practice (within some degree of confidence), and give instant feedback via various prompting techniques.

Figure 1. Views of the simulated classroom and virtual student created in Unity. Users would be immersed in this room and verbally interact with the student within the VR headset.

(a) View facing the avatar (b) View of the Classroom

The initial pilot study [J1] consisted of a multiple-baseline across participants and behaviors single-case design. Lessons concerned the use of a nondirective mathematical questioning strategy in instances where a simulated student provided correct or incorrect answers to word problems. This involved staggering the implementation of the intervention across two participants and lessons over the course of one month. Measures were observed and automated assessments of participant performance and subjective assessments. The experiment incorporated changes needed to improve simulation presentation and performance. A subsequent evaluation of a refined, feature-locked version of the simulation was then conducted [J2]. We recruited and randomly assigned 30 college-aged participants into treatment (i.e., lecture, modeling, VR) and control groups (i.e., lecture and modeling only). Participants completed pre-test, post-test, and maintenance probes each week over the course of three weeks. Lessons concerned the use of a nondirective mathematical questioning strategy in instances where a simulated student provided correct or incorrect answers to word problems. Measures included observed and automated assessments of participant performance and subjective assessments of participant confidence.

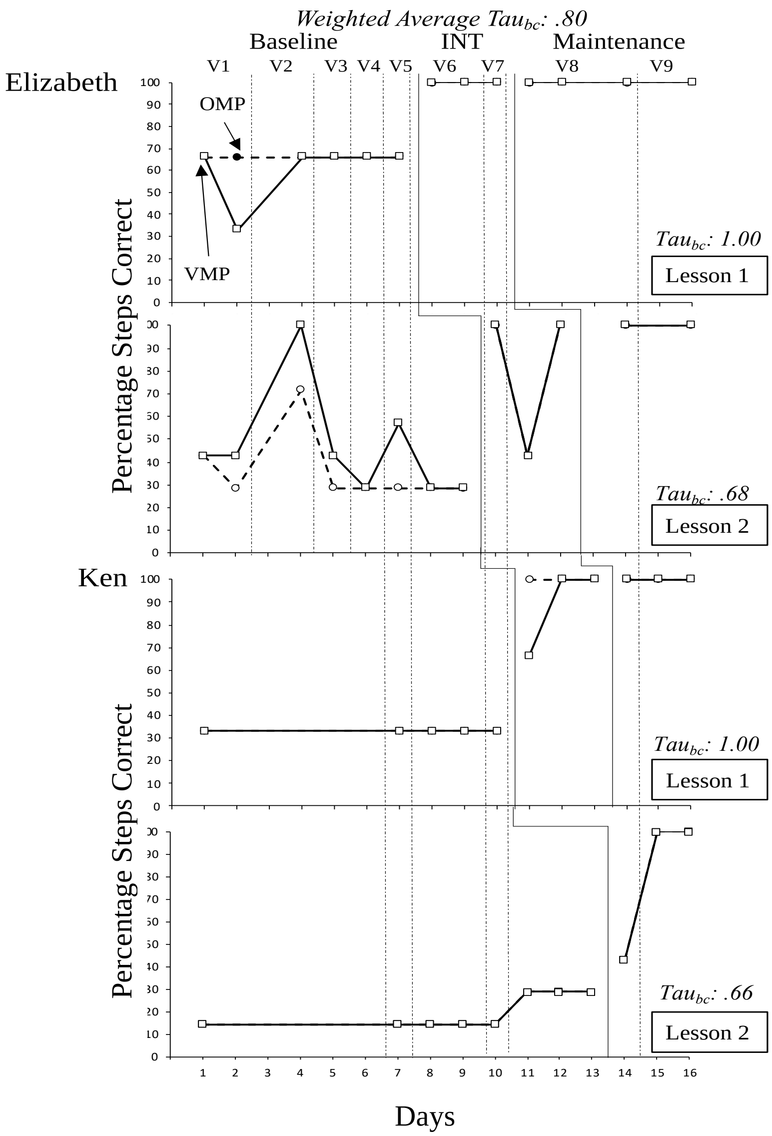
Results:

For King et al. [J1], we demonstrated the effectiveness of training consisting of video-recorded lectures and smart simulations capable of automatic assessment, prompting, and feedback through the incorporation of VR and AI (e.g., speech classification, speech-to-text) to improve the use of an interactive, speech-based academic teaching strategy. Use of single-case design permitted improvements to simulation functionality, resulting in automated assessments with a high degree of agreement with direct observation (>96%) and large changes in the percentage of steps in the procedure exhibited by participants ($n = 2$; $Tau_{bc} = .80$; See Figure 2 below; see Tarlow, 2017 for an explanation of effects).

A subsequent randomized control trial [J2] of a refined version of the simulation revealed robust improvements in the ability of the treatment group ($n = 15$) to deliver the procedure, relative to a control group ($n = 15$) who received didactic instruction. A mixed ANOVA revealed significant main effects of time ($F[2,27]=124.154$, $P < .001$, $\eta_p^2 = 0.816$) and treatment ($F[1,28]=19.281$, $P < .001$, $\eta_p^2 = 0.408$), as well as an interaction effect ($F[2,28]=8.429$, $P < 0.001$, $\eta_p^2 = 0.231$) for the average percentage of steps in the questioning procedure (See Figure 3). Post-test scores for the intervention group ($M = 88\%$, $SD = 22.62$) exceeded control group performance ($M = 63.33\%$, $SD = 22.64$), with $t[28] = 3.653$, $P < .001$, $d = 1.334$. Maintenance scores indicated a positive effect of intervention ($M = 83.33\%$, $SD = 24.40$) relative to control ($M = 54.67\%$, $SD = 15.98$), $t[28] = 3.807$, $P < .001$, $d = 1.39$. A Mann-Whitney U test indicated the treatment groups' self-ratings of confidence ($M = 2.41$, $SD =.51$) were higher than those of the control group ($M = 2.04$, $SD =.52$), $U = 64$, $P = .043$, $r = .137$).

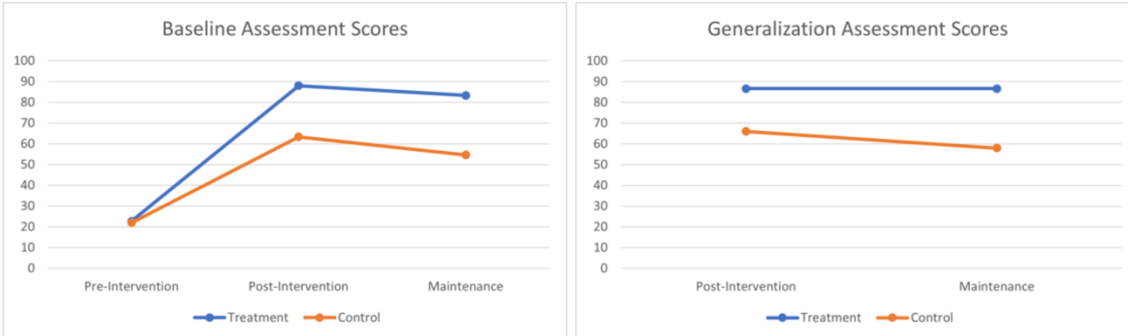
Overall, these two studies have demonstrated that a 'smart' VR training simulation can be used to realize improvements in conversation-based skill acquisition, maintenance, and perceived levels of confidence. While this has initially been demonstrated within the context of mathematical questioning strategies, future work will aim to incorporate additional features into the simulation (e.g., motion capture animation, assessment of trainee motion) to address a broader range of practices in education and other disciplines.

Figure 2. Percentage of Questioning Procedure Steps Completed Correctly During Pilot Study



Note. Open squares with solid lines depict virtual mastery probes (VMP). Open circles with the dashed line represent observed mastery probes (OMP). Overlap of data paths indicates 100% correspondence. The solid circle indicates a misadministration. Intervention sessions (INT) consisted of different components based on specific days, with didactic instruction provided on the first of intervention, error-free prompts provided on the first day and every subsequent session in which participants did not meet a criterion of 80% steps correct, and video models and two delayed prompting sessions provided every day of the intervention. Performance under prompted conditions not displayed. Changes associated with different versions included: V2 changed the classifier and improved head tracking of simulated student; V3 increased the confidence threshold of classifier to .5; V4 added simulated student speech, changes to content of student comments, and decreased latency; V5 removed onscreen student response text; V6 added phrases to the language classifier and increased the confidence of the classifier to .75, and altered student text; V7 provided minor cosmetic changes; V8-9 added items to the classifier. INT= intervention; V1-9 = simulation version 1-9.

Figure 3. Percentage of Questioning Procedure Steps Completed by Treatment and Control Groups



(a) Baseline assessment scores

(b) Generalization assessment scores

Ideas/aims for future extramural project:

We aim to extend our success in a primarily verbal procedure into areas in which additional modalities are required for success. Specifically, we will (1) modify the teaching platform to skills aligned with the functional behavior assessment, a procedure commonly used to address the problem behaviors of students with disabilities and (2) evaluate the feasibility of training practitioners using the automated VR platform.

Publications resulting from project:

Publications

[J1] King, S.A., Estapa, A., Bell, T., & Boyer, J. "Behavioral skills training through smart virtual reality: Demonstration of feasibility for a verbal mathematical questioning strategy." *Manuscript under review.*

[J2] King, S.A., Boyer, J., Bell, T., & Estapa, A. "A smart virtual reality training system for teacher-student interaction." *Manuscript in preparation.*

Internal Grants

King, S.A., Estapa, A., & Bell, T. (2020). Immersive Virtual Mathematics Education for Teachers: Project VIME. ($18,000) Obermann Center Interdisciplinary Research Grant. Funded.

King, S. A., & Bell, T. (2021). Immersive Virtual Instruction in Behavioral Education: Project VIBE. Innovations with Teaching Technology Award. ($56,180). Funded.

King, S.A., Bell, T., Higgins, W., & Gehringer, J. (2022). The Virtual Immersive Training Project in Special Education Project. Iowa Center for School Mental Health ($250,000). Under review.

External Grants

King, S. A., Estapa, A., Bell., T, & Higgins, W. (2021). Virtual Immersive Simulation Training in Mathematics and Special Education. ($849,650). National Science Foundation. Not funded.

Theses

Boyer, J. (2022). Customizable Smart Virtual Reality Simulations for Conversation-based Training. M.S. in Electrical and Computer Engineering.