Just *Look* at Them! Encouraging Self-Reflection on Teacher Gaze Behavior through Data Visualizations in Virtual Reality

Yejoon Yoo*
Cornell University
Ithaca, NY, USA
yy937@cornell.edu

Jonathan Segal*
Cornell University
Ithaca, NY, USA
jis62@cornell.edu

Aleshia Taylor Hayes
University of North Texas
Denton, TX, USA
aleshia.prof@gmail.com

Andrea Stevenson Won
Cornell University
Ithaca, NY, USA
asw248@cornell.edu

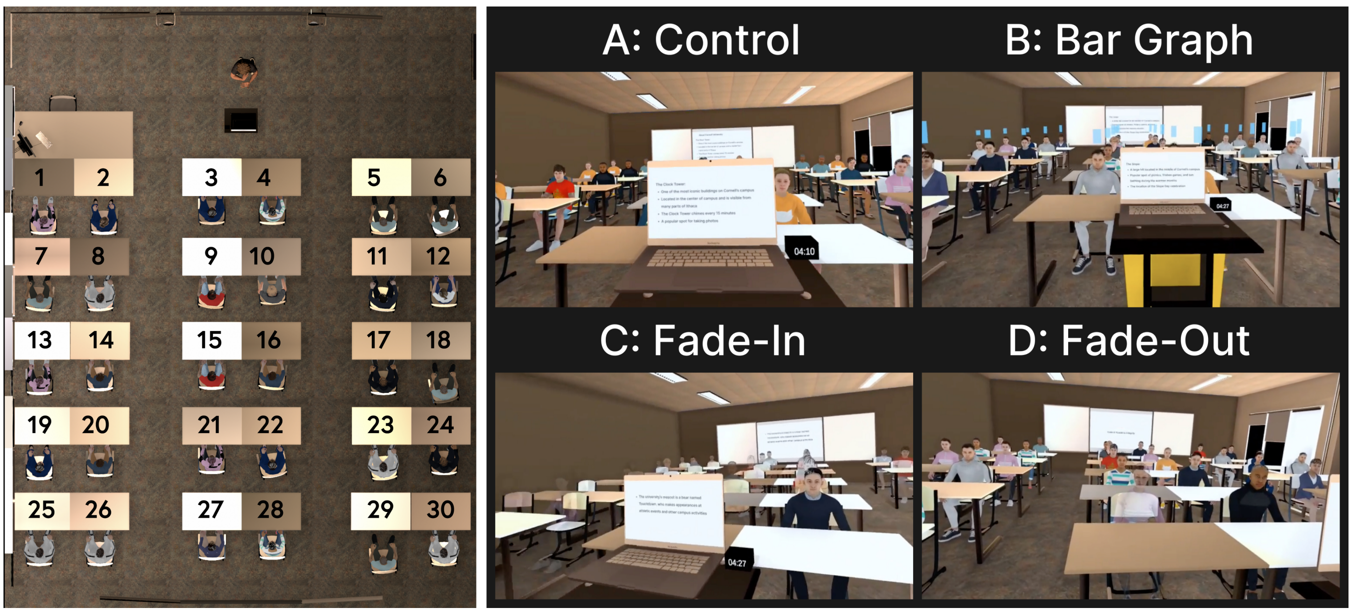

Figure 1: On the left, a bird's-eye view of the virtual classroom; on the right, screenshots of each of the four gaze-visualization conditions

ABSTRACT

Virtual classroom simulations are an exciting avenue to give teachers a way to reflect on their teaching behavior, in particular, nonverbal behavior. In two within-participants studies, we explore how visualizing participants' gaze, using four different data visualizations, affected participants' behaviors and self-reflection in an immersive virtual reality classroom simulation. We compared a Control condition with no data visualization, an updating Bar Graph over each "student" agent head, and two Fade-In/Fade-Out conditions where the opacity of "students" changed based on whether they were in the field of view of the participant. We found that participants preferred the Bar Graph visualizations, and this condition changed participants' behavior the most compared to the Control condition. We discuss design implications for virtual classroom simulations as a self-reflection tool for teachers.

*Both authors contributed equally to this research.

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3651105

CCS CONCEPTS

• **Human-centered computing** → **Visualization systems and tools**; **Empirical studies in interaction design**; **Empirical studies in HCI**; **Empirical studies in visualization**.

KEYWORDS

3D Data Visualization, User Interface, Immersive Visualization, Gaze Data, Virtual Reality Simulation

ACM Reference Format:
Yejoon Yoo, Jonathan Segal, Aleshia Taylor Hayes, and Andrea Stevenson Won. 2024. Just *Look* at Them! Encouraging Self-Reflection on Teacher Gaze Behavior through Data Visualizations in Virtual Reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3613905.3651105

1 INTRODUCTION AND RELATED WORK

Virtual reality (VR) simulations allow users to practice interactions in a replication of the physical environment using naturalistic behaviors [5, 22]. Some of the earliest uses of virtual reality simulations include improving teaching practice through virtual classroom simulations that allow teachers to observe and analyze their own teaching behaviors based on real-time feedback [3, 16]. A better understanding of how virtual reality can help people transform nonverbal behavior, including gaze behavior, has broad utility for classroom simulators [3], public-speaking practice applications, and any other virtual environment in which participants might wish to monitor and modify their behavior [18, 27].

A key aspect of classroom management is the teacher's ability to distribute their gaze (and their associated attention) throughout the classroom. In this paper, we describe two within-participant pilot studies investigating the effectiveness of different *types* of gaze data visualization in changing participants' behavior in a virtual classroom. We compared four conditions: 1. control, with no gaze data visualization; 2. gaze data represented through Bar Graphs; 3. gaze data presented through the opacity of students, following [3] (Fade-In); and 4. an opposing, Fade-Out condition with gaze data presented through the inverted opacity of the students. In Study 1, we used head movement as a proxy for gaze. In Study 2, we used eye tracking directly. We aimed to explore approaches to provide real-time visualizations of teacher gaze behavior in a classroom simulation, and examine behavior changes, usability, and user experience.

Below, we review related works on the utility of simulations for teaching and popular classroom simulators. We discuss the concept of transformed social interaction, specifically in the context of teaching simulations designed to promote self-reflection focusing on nonverbal behavior change. We discuss our study results, and close by discussing next steps for teaching simulators using transformed social interaction.

1.1 Classroom Simulations

Given increasing rates of teacher attrition, effective teacher preparation has become increasingly important in recent years. This preparation for teachers has ranged from role-play in classrooms, to practice teaching, to computer-based classroom experience simulations [26]. Teaching simulations provide new methods of visualization and presentation [21] allowing teachers to practice skills and learn from real-time feedback in ways not possible in real life [4]. Such systems have taken various forms, including highly immersive, interactive 3D virtual environments [24] and other laboratory-based setups that incorporate motion tracking and large displays (e.g., TeachLivE [10, 15]). These virtual classroom simulators allow teachers and teachers in training to practice, improving target skills and general teaching skills in a controlled simulation, in ways that transfer to the classroom where they engage with "real" students [14].

Cues such as eye gaze [1, 6, 20] are particularly important in the classroom. Eye contact engages students, encourages participation, manages classroom behaviors, and creates a sense of psychological closeness and connection between teachers and students [8, 12, 25]. In this study, we used gaze behaviors as a way of investigating whether more abstract or more literal representations of nonverbal behavior are more useful in leading users of a teaching simulator to reflect on their behavior. As a starting point, we asked students to make an effort to deploy their gaze evenly, as literature has identified even gaze distribution as an indicator of expert teachers [7, 9]

1.2 Transformed Social Interaction as a Teaching Tool

Virtual reality offers the unique ability to "transform social interaction" [2] in ways that could potentially improve on face-to-face interaction Immersive virtual reality allows tracked behavior to be modified to improve the outcomes of an interaction, for example, by replacing unwanted behaviors (staring at one's laptop, nervous gestures) with desirable behaviors (appropriate eye contact, good posture). While some versions of transformed social interaction focus on allowing a person to control how their behavior appears to their *audience*[3], these transformations can also be made available to the user, so they can become more aware of, and potentially modify, their own behavior. For example, in a 2008 study, researchers created a virtual reality classroom simulator that linked the appearance of the "students" to the gaze behavior of the "teacher" participant. The student agent-avatars would slowly become transparent **(Fade-In)** if the teacher participant didn't look at them, leading the "teachers" to distribute their gaze more evenly. Our first research question aims to replicate this result: *RQ1: Can transforming how participants' gaze is represented during a virtual teaching task affect how participants distribute their gaze around a virtual classroom?*

However, participants might not even notice vanishing students if they were too focused on the middle of the classroom. Alternatively, students could become transparent **(Fade-Out)** as the teacher gazes at them, reminding teachers to look at the rest of the virtual classroom. Thus, *RQ2a: which is more effective in leading participants to distribute their gaze evenly: a metaphor of neglect (leading participants to restore attention to students who are fading away) or a metaphor of over-attention (where students receiving a disproportionate percentage of gaze slowly disappear from view)?*

In addition, because high cognitive load is known to leading teachers to rely on biases and reflexive behaviors [11] and has been shown to reduce the efficacy of teacher training interventions [17], it is important to monitor the level of cognitive load users are experiencing in these training tools. Thus, *RQ2b: which is perceived as more cognitively difficult: a metaphor of neglect, or a metaphor of overattention?*

Both representations of gaze involving transparency (Fade-In and Fade-Out) draw on very visual metaphors of attention, but are somewhat abstract. Thus, we created a third measure, in which a **Bar Graph** representing the average gaze distributed to each virtual student appears over that student-avatar's head, leading to these research questions: *RQ3a: Is a more explicit measure more effective than a more implicit measure in changing participants' gaze patterns? RQ3b: Is a more explicit measure perceived as more effective than a more implicit measure?* and *RQ3c: Is a more explicit measure perceived as more cognitively difficult than a more implicit measure?*

Finally, we investigated whether repeated simulation experiences in a virtual classroom improve participants' gaze behavior over time: *RQ4: Does gaze behavior become more evenly distributed as participants complete the four conditions?*

2 STUDY 1

In Study 1, participants assumed the role of a teacher in a virtual classroom setting lecturing to thirty virtual "students" (agent-avatars). Each participant completed all four conditions (shown in Figure 1) in a randomly determined order. In the Control condition, participants saw no data visualization. In the Bar Graph condition, bar graphs over each student's head grew higher when the teacher directed their gaze at a particular student and lowered when they looked elsewhere. The Fade-In and Fade-Out conditions represented gaze behavior via the opacity of each student. Students "fade-in" when the participant looks at them and slowly disappear when the participant is not looking at them (following [3]). The Fade-Out condition reverses this: students become more transparent as the participant gazes at them.

2.0.1 Participants. For Study 1, forty-two students (26 female, 15 male, and 1 non-binary) participated, compensated with course credit or $25. Six participants were excluded due to data loss, and eight due to technical issues. All participants signed informed consent, and all procedures were approved by the Institutional Review Board (IRB).

2.0.2 Materials. The virtual classroom was developed in Unity 3D, using pre-existing and custom-built assets. The Oculus Unity Plugin was used for HMD integration. "Student" agent-avatars were imported from Mixamo via their plugin application. The position of the students was randomized for each session. In addition to an idle sitting animation from Mixamo.com, a script animated the students' heads to orient towards the participant. Other than this head-direction interaction, no other interaction was scripted for the students. A laptop on a podium in front of the classroom contained pre-prepared slides for four lectures on university-related topics. Slides for the Control condition covered "Traditions". In the Bar Graph condition, slides discussed "Popular Destinations"; in the Fade-In condition, "Fun Facts." Finally, the Fade-Out condition slides described the "Academic Integrity Code." Participants were instructed to lecture to the students based on the presented slides.

Participants wore an Oculus Quest 2 and controlled the slides using the Quest controllers. In Study 1, participants controlled a generic blocky orange avatar, whose hands and bodies they could see if they looked straight down. Each trial was limited to 5 minutes. After five minutes, a pop-up message instructed participants to remove the headset.

We detected participant gaze as follows. "Direct gaze" was calculated when the participant's raycast directly contacted a given student. "Peripheral gaze" was defined as occurring when a cone representing the field of view of the Meta Quest 2 (134 degrees, [19]) collided with a given student avatar.. There were three states for each "student". Direct gaze counted as "1", peripheral gaze as 0.5", and no gaze was counted as "0". These values were added to a circular queue for each student 20Hz. The queue was set to 200 to represent the last 10 seconds of gaze information. This value was used to determine the visualizations for each student in each condition. The measure of "total gaze" was calculated as a rolling average of two separate values over ten seconds. which drove the changes in all three visualizations. The authors created a custom shader that maintained the visual integrity of the students while allowing them to be transparent at a variable percentage.

2.0.3 Procedure. On arrival, participants received an overview of the study, asked any questions, and signed informed consent. Participants were told they would present four different pre-prepared lectures in English to virtual students, and were instructed to try to spread their eye gaze equally between all 30 students. Participants were then asked to adjust the HMD and move around the VR classroom to get comfortable with the controls and environment. After five minutes in each condition, participants removed their headsets and completed the NASA-TLX survey. After completing all four conditions they completed a post-test questionnaire. The entire experiment took about 40 minutes.

2.0.4 Measures. We collected two types of data; participants' tracked *behavioral data* in the virtual classroom, and *self-report data*, based on participant responses after each condition and at the end of the study.

Behavioral Data. The position and state data for the "student" agent-avatars, slides, clock, and the HMD and hand controllers were recorded 20 times a second and saved to a file on the device. Researchers replayed each file to observe the experiment from the participant's point of view and ensure that there were no unreported issues that would require exclusion. Based on this review, we used only the first three minutes of recording, since many participants finished the slides more quickly than five minutes and their behavior was then not typical of teaching. To represent **Gaze Attention Time**, we generated 'heat maps" to represent participants' direct gaze by mapping the seconds of gaze each participant directed towards each "student" over the entire three-minute interaction.

Self-Report Data. The NASA Task Load Index (NASA-TLX) measures perceived workload or cognitive load experienced by individuals during a task [13] in six dimensions Mental, Physical, and Temporal Demand, Performance, Effort, and Frustration. Participants rated each of these aspects from 0 ("very low") to 100 ("very high"). We excluded Performance scores since a number of participants were confused about this scale.

At the end of the study, participants assessed the effectiveness of the four conditions using 7-point Likert scales, and responded to six open-ended questions about their preferences and overall experience (See Appendix for full questions).

2.1 Study 1 Results

We present an analysis using visualizations, statistical tests, and qualitative analysis conducted in R [23].

2.1.1 Study 1 - Behavioral Data. Our first research question sought to replicate the initial findings of [3] *RQ1: Can transforming how participants' gaze is represented during a virtual teaching task affect how participants distribute their gaze around a virtual classroom?* We calculated the time that gaze behavior was directed to each of the 30 "students" during the initial three minutes of teaching in each condition. Using these measures of total direct gaze per student as our dependent variable, in an linear mixed-effects model using condition and order as fixed effects, we found that compared to the control condition, all three visualization conditions increased the time that participants looked at students $F(3,9709)=7.12$, $p < .001$. However, addressing *RQ2a: Is a metaphor of neglect (Fade In) more effective than a metaphor of over-attention (Fade-Out)?* and *RQ3a: Is a more explicit measure more effective than a more implicit measure in changing participants' gaze patterns?*; we found no statistically significant differences overall *between* the three visualization conditions (all p's > .10). To visualize these patterns more granularly, we generated heatmaps for each condition, representing the distribution of direct gaze behavior across students, shown in Figure 2. Addressing RQ4, figure 3 shows heatmaps by order; however, we did not see consistent effects of time.

2.1.2 Study 1 - Self-report. To address our remaining research questions, we turned to the self-report data. The Bar Graph visualization consistently produced the lowest scores across all five dimensions, including mental demand, addressing *RQ2b: Is a metaphor of neglect perceived as more cognitively difficult than a metaphor of overattention?* and *RQ3c: Is a more explicit measure perceived as more cognitively difficult than a more implicit measure?* However, using a linear mized-effects model with condition and order as fixed effects, participants did not rate conditions significantly differently on any of the NASA-TLX measures. Because participants were confused about the performance measure, we did not address *RQ3b: Is a more explicit measure* ***perceived*** *as more effective than a more implicit measure.*

In line with the NASA-TLX results, participants expressed a higher preference for Condition B (Bar Graph) and Condition C (Fade In), while displaying lower preferences for Condition C and Condition D (Fade-Out) (see Figure 5).

Themes emerging from the analysis of user comments were more frequently positive for the Bar Graph condition. Negative themes (e.g. confusing, stressful) were expressed more and in greater magnitude towards the Fade-In and Fade-Out conditions. While only one participant was confused by with the bar charts, eight were confused by the Fade-In / Fade-Out visualizations. It was also noteworthy that only one participant indicated that they did not understand what the data visualizations represented. Also reassuringly, given that we used head position and orientation as a proxy for gaze, participants regularly described the visualizations as representing "gaze" or "eye contact" with the students. Other salient themes emerged regarding user experience. Negative aspects of user experience included discomfort with the hardware (e.g. weight, fit, difficulty with glasses) as well as the interface (blurriness and motion sickness). User satisfaction was coded in positive comments about novelty, fun, and the sense of presence, immersion, and realism.

2.2 Limitations of Study 1

In this small exploratory study, we were limited to a population of convenience whose teaching experiences were primarily limited to serving as a teaching assistant. A perhaps more serious limitation was our use of head position as a proxy for gaze. While this has been used in previous work [3], eye trackers, which are increasingly available in VR headsets, can provide more precise measurements of gaze attention time compared to head orientation. In addition, participants were not given the option to customize their avatars, the lecture topics were confounded with conditions, and participants went through the lecture materials more quickly (three minutes) than expected from pre-testing. Given these issues, we conducted a second pilot study to address these issues and replicate Study One's findings.

3 STUDY 2

3.1 Methodological Changes for Study 2

In Study Two, our measures, analyses and IRB approval were identical to Study 1. However, we switched from the Meta Quest 2 to the Meta Quest Pro headset to implement eye tracking. We used the Meta Unity SDK to attach eye tracking to meshes representing the eyeballs, and used these meshes as raycast origins to determine which students were targets of participants' gaze. Because this directly captured gaze, we did not simulate peripheral gaze in Study 2.

In addition, we removed the participant's first-person avatar entirely to avoid potential confounds of customization, instead creating transparent ghost hands for the controllers. We also removed the clock on the left wall and lowered the height of the podium. Finally, we counterbalanced both the order of teaching materials and conditions. Each participant was provided with two randomly selected sets of slides to support a five-minute lecture for each condition. Slides covered eight topics: 'Social Media's Impact on Society,' 'Cybersecurity,' 'Driving Manners and Road Safety,' 'Public Transportation,' 'Diversity,' 'Sexual Misconduct,' 'Alcohol and Other Drugs,' and 'Health and Wellness.' We also provided additional instructions about rating the performance category of NASA-TLX, as participants were previously confused about the perfect score being 0 in Study 1. The procedure for Study 2 was otherwise identical to Study 1.

3.1.1 Participants. Twenty-eight students (20 female and 8 male) participated in Study 2, choosing between course credit or $25. Four participants were excluded due to technical issues, and two due to a delayed start.Participants in Study 1 were excluded from Study 2. .

3.2 Study 2 - Results

3.2.1 Study 2 - Behavioral Data. We again calculated the time that gaze behavior was directed to each of the 30 students. After expanding the lecture materials, we could use the initial *four* minutes of teaching in each condition. Repeating the linear mixed-effects model from Study 1, we found two of the three visualization conditions (Bar Graph, and Fade-In) increased the time that participants looked at students compared to control $F(3,9828)=9.85$, $p < .001$. In addition, using the emmeans package for pairwise comparison,

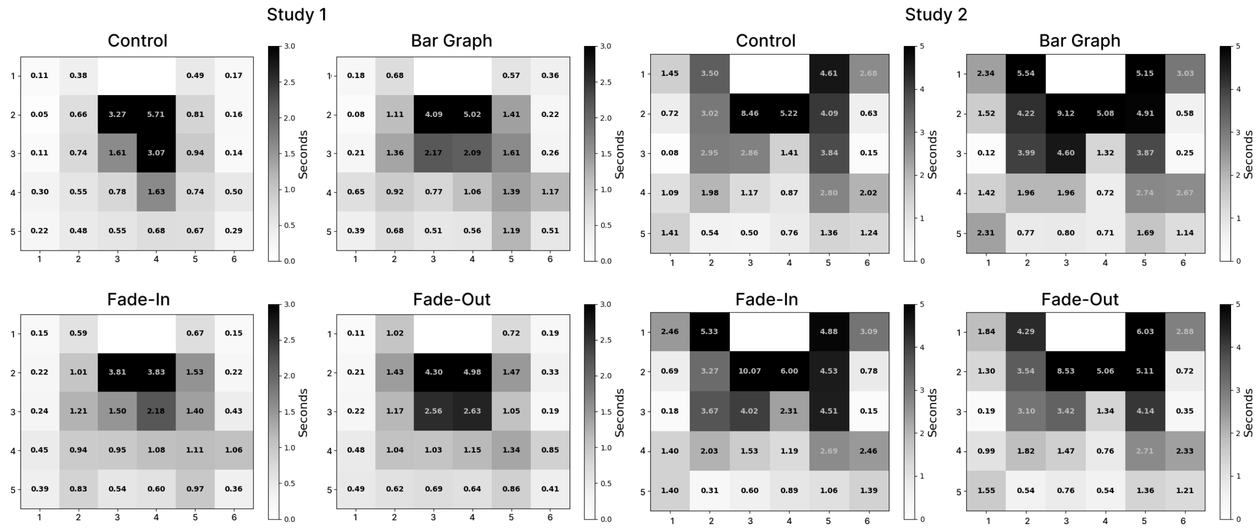

Figure 2: The direct gaze heatmaps of Study 1 and Study 2 depict the average total direct gaze attention time across a total of three minutes, measured in seconds, for thirty student agents organized in five rows of desks across four distinct conditions. The conditions with visualizations present more evenly distributed direct gaze attention compared to the control condition. We excluded values from Col 3 and Col 4 of Row 1 in all heatmaps, as participants' view of the students was confounded by the podium.

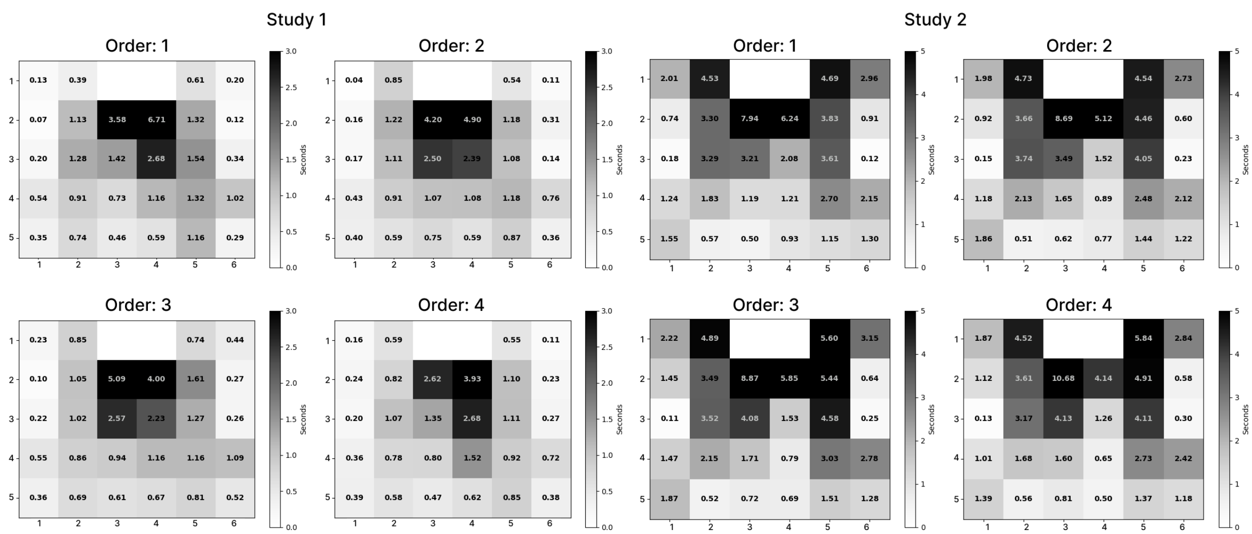

Figure 3: The heatmaps of mean direct gaze by the order of Study 1 and Study 2 conditions reveals that variations in the order of conditions had little impact on the distribution of head orientation associated with gaze attention. We excluded values from Columns 3 and 4 of Row 1 in all heatmaps, as the student agents in these positions were sometimes in line of the participants' peripheral gaze when they were looking at the teaching materials on the podium.

we found that the Bar Graph condition statistically significantly increased gaze over the Fade-Out condition (p = .020). Figure 2 reflects these findings.

For Study 1, the gaze depicted in Figure 2. was calculated using head direction. In study 2, we used eye tracking.

Comparing Study 1 and Study 2 (Figure 2), we see in both, participants' gaze was more evenly distributed in the Bar Graph conditions.

3.2.2 Study 1 and Study 2 - Self-Report. Participants rated the overall experience in Study 2 higher in every measure compared to

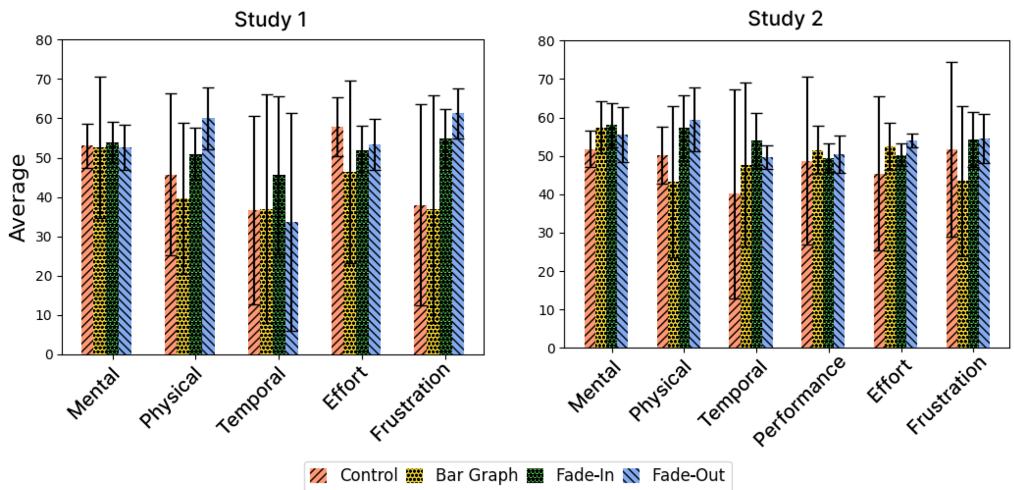

Figure 4: Study 1 and Study 2 graphs of NASA-TLX results.

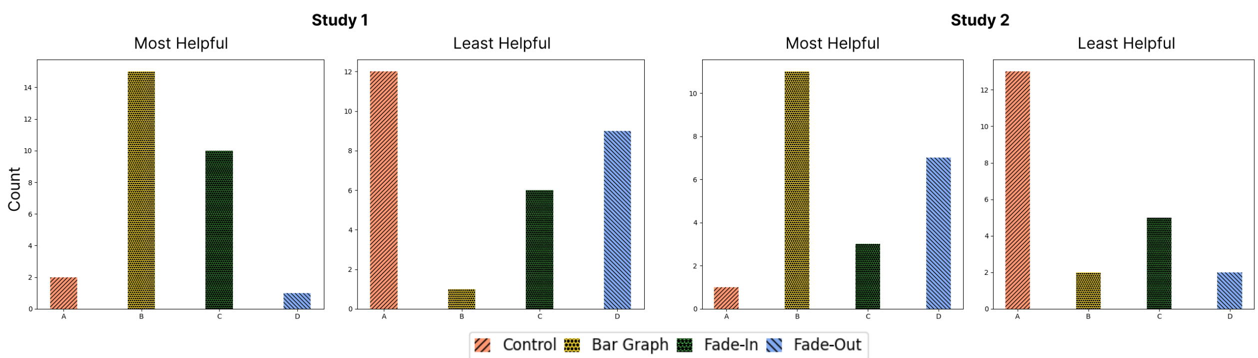

Figure 5: Study 1 and Study 2 graphs of participants' preferences on conditions. The participants were asked which conditions were most and least helpful in improving their gaze behaviors during the experiments.

Study 1. Participants' understanding of gaze data visualization increased from 5.73 to 6.31 in the post-test questionnaire. Participants also expressed their intention to use the simulation to enhance their nonverbal teaching behavior, as indicated by an increase in values from 5.83 to 6.13.

The NASA-TLX results from Study 2 are almost identical to those of Study 1, presenting higher cognitive load in the visualization conditions compared to control. (After providing additional instructions to participants, we were able to include the Performance score). As before, we did not see statistically significant differences between conditions on any TLX measures (all p's > .10) Participants' preferences differed slightly between Study 1 and Study 2, with an increased preference for the Fade-Out condition, but continued to prefer the Bar Graph condition overall (Figure 5).

3.3 Discussion

In two within-participant studies, we find convergent evidence supporting the effectiveness of visualizations of gaze behavior in helping participants distribute their gaze more evenly within a virtual classroom. Participants demonstrated distinct behavioral patterns depending on the condition. Analyzing survey and qualitative responses, we found that participants preferred a more explicit (Bar Graph) representation of gaze. Study 2, which used eye-tracking to capture gaze, led to a more even distribution of gaze than Study 1, which used head movements as a proxy. However, even in Study 1, participants regularly described the visualizations as representing "gaze" or "eye contact" with the students.

Some of the studies' limitations also indicate next steps. Our study focused on a classroom simulator that could be used to practice techniques that could then be employed in a real classroom. We instructed participants to try to distribute their gaze equally. However, in a real classroom, optimal gaze may not be equal over time, but rather determined by timing and in reaction to student movements. Similarly, if a teacher was actually teaching on a daily basis in a virtual classroom, more implicit measures might be more acceptable; however, a virtual classroom would allow the layout to be redesigned, as well as other affordances only possible in virtual environments.

To further explore different behavioral patterns associated with different visualizations and interactions, future work should compare teaching performance in the VR classroom simulator to teaching in a real classroom environment (whether virtual or physical). Future research could also include augmented or mixed reality applications to offer feedback on teaching performance while instructors are actively teaching in a physical classroom setting.

In addition to providing observations on the behavioral patterns associated with gaze attention based on visualization of real-time feedback, participants proposed other modifications that might be valuable for developing virtual reality classroom applications.

- *The implementation of new classroom layouts in VR simulation could offer a unique opportunity to practice various settings and scenarios.*
- *Enhancements to the Bar Graph simulation could display cumulative gaze attention data, enabling participants to monitor their performance over time*
- *More realistic reactions from student agents, especially nonverbal behavior, would enhance the experience.*

4 CONCLUSION

Decisions to implement VR technology in an educational context must weigh the advantages of novel interactions against the cognitive load of novel interfaces. In this paper, participants found value in virtual reality tools to aid self-reflection in a virtual classroom simulator despite cognitive load. Technological improvements (eye tracking instead of head movements as a proxy for gaze) also improved participants' ratings of the experience. Whether in a real classroom or in a training simulator, updated tools will continue to offer these challenges and opportunities.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2234802

The authors would like to thank Krupa Sekhar, Yohanes Kidane, and, Samuel Rodriguez for their work on the development of the project.

REFERENCES

[1] Michael Argyle and Mark Cook. 1976. *Gaze and mutual gaze.* Cambridge U Press, Oxford, England. Pages: xi, 210.

[2] Jeremy N Bailenson, Andrew C Beall, Jack Loomis, Jim Blascovich, and Matthew Turk. 2004. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators & Virtual Environments* 13, 4 (2004), 428–441.

[3] Jeremy N. Bailenson, Nick Yee, Jim Blascovich, Andrew C. Beall, Nicole Lundblad, and Michael Jin. 2008. The Use of Immersive Virtual Reality in the Learning Sciences: Digital Transformations of Teachers, Students, and Social Context. *Journal of the Learning Sciences* 17, 1 (Feb. 2008), 102–141. https://doi.org/10.1080/10508400701793141

[4] Oluleke Bamodu and Xuming Ye. 2013. Virtual Reality and Virtual Reality System Components. *Advanced Materials Research* 765-767, 978-94-91216-42-8 (April 2013). https://doi.org/10.2991/icsem.2013.192

[5] Grigore C. Burdea and Philippe Coiffet. 2003. *Virtual reality technology (2. ed.).* Wiley.

[6] Judee K Burgoon, Michael Pfau, Thomas Birk, and Valerie Manusov. 1987. Nonverbal communication performance and perceptions associated with reticence: Replications and classroom implications. *Communication Education* 36, 2 (1987), 119–130.

[7] Muhammad Butt. 2011. Eye Contact as an Efficient Non-verbal Teaching Technique: A Survey of Teachers' Opinion. *European Journal of Social Sciences* (Jan. 2011).

[8] Muhammad Naeem Butt and Mohammad Iqbal. 2011. Teachers Perception Regarding Facial Expressions As An Effective Teaching Tool. *Contemporary Issues in Education Research (CIER)* 4, 2 (March 2011), 11–14. https://doi.org/10.19030/cier.v4i2.4077

[9] Kai S Cortina, Kevin F Miller, Ryan McKenzie, and Alanna Epstein. 2015. Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. *International Journal of Science and Mathematics Education* 13 (2015), 389–403.

[10] Melanie Rees Dawson and Benjamin Lignugaris/Kraft. 2017. Meaningful practice: Generalizing foundation teaching skills from TLE TeachLivE™ to the classroom. *Teacher Education and Special Education* 40, 1 (2017), 26–50.

[11] David F Feldon. 2007. Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational psychologist* 42, 3 (2007), 123–137.

[12] Carolynn C. Hamlet, Saul Axelrod, and Steven Kuerschner. 1984. Eye contact as an antecedent to compliant behavior. *Journal of Applied Behavior Analysis* 17, 4 (1984), 553–557. https://doi.org/10.1901/jaba.1984.17-553

[13] S.G. Hart and L.E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52 (1988), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[14] Aleshia T Hayes, Stacey E Hardin, and Charles E Hughes. 2013. Perceived presence's role on learning outcomes in a mixed reality classroom of simulated students. In *Virtual, Augmented and Mixed Reality. Systems and Applications: 5th International Conference, VAMR 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part II 5.* Springer, 142–151.

[15] Aleshia T Hayes, Carrie L Straub, Lisa A Dieker, Charlie E Hughes, and Michael C Hynes. 2013. Ludic learning: Exploration of TLE TeachLivE™ and effective teacher training. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)* 5, 2 (2013), 20–33.

[16] Crystal L. Hoyt, Christopher Rex, and Kimberly R. Swinth. 2003. Social Inhibition in Immersive Virtual Environments. *Presence Teleoperators Virtual Environ.* 12, 2 (2003), 183–195. https://doi.org/10.1162/105474603321640932

[17] Yizhen Huang, Eric Richter, Thilo Kleickmann, Axel Wiepke, and Dirk Richter. 2021. Classroom complexity affects student teachers' behavior in a VR classroom. *Computers & Education* 163 (2021), 104100.

[18] Isabel L. Kampmann, Paul M. G. Emmelkamp, Dwi Hartanto, Willem-Paul Brinkman, Bonne J. H. Zijlstra, and Nexhmedin Morina. 2016. Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial. *Behaviour Research and Therapy* 77 (2016), 147–156. https://doi.org/10.1016/j.brat.2015.12.016

[19] Jonathan W. Kelly, Taylor A. Doty, Morgan Ambourn, and Lucia A. Cherep. 2022. Distance Perception in the Oculus Quest and Oculus Quest 2. *Frontiers in Virtual Reality* 3 (2022). https://doi.org/10.3389/frvir.2022.850471

[20] Chris L Kleinke. 1986. Gaze and eye contact: a research review. *Psychological bulletin* 100, 1 (1986), 78.

[21] Veronica S. Pantelidis. [n. d.]. Reasons to Use Virtual Reality in Education and Training Courses and a Model to Determine When to Use Virtual Reality. *Themes in Science and Technology Education* v2 n1-2, 2009 ([n. d.]), p59–70.

[22] Shiva Pedram, Stephen A. Palmisano, Richard Skarbez, Pascal Perez, and Matthew Farrelly. 2020. Investigating the process of mine rescuers' safety training with immersive virtual reality: A structural equation modelling approach. *Comput. Educ.* 153 (2020), 103891. https://doi.org/10.1016/j.compedu.2020.103891

[23] R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[24] Kalliopi-Evangelia Stavroulia and Andreas Lanitis. 2023. The role of perspective-taking on empowering the empathetic behavior of educators in VR-based training sessions: An experimental evaluation. *Comput. Educ.* 197 (2023), 104739. https://doi.org/10.1016/j.compedu.2023.104739

[25] Fiona D. H. Tan, Peter R. Whipp, Marylène Gagné, and Niels Van Quaquebeke. 2019. Students' perception of teachers' two-way feedback interactions that impact learning. *Social Psychology of Education* 22, 1 (Feb. 2019), 169–187. https://doi.org/10.1007/s11218-018-9473-7

[26] Xu Wang, Meredith Thompson, Kexin Yang, Dan Roy, Kenneth R. Koedinger, Carolyn P. Rose, and Justin Reich. 2021. Practice-Based Teacher Questioning Strategy Training with ELK: A Role-Playing Simulation for Eliciting Learner Knowledge. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–27. https://doi.org/10.1145/3449125

[27] Hangyu Zhou, Yuichiro Fujimoto, Kanbara Masayuki, and Kato Hirokazu. 2021. Virtual Reality as a Reflection Technique for Public Speaking Training. *Applied Sciences* 11, no. 9: 3988 (2021).

A APPENDIX: STUDY 1 PARTICIPANTS

Table 1 presents information on Study 1 participants, encompassing demographics, recruitment, dropouts, and the total number.

Category	Recruitment	Dropout	n
Gender			
Female	26	6	20
Male	15	7	8
Non-binary	1	1	0
Race			
Asian	30	10	20
Caucasian	6	2	4
Multi-racial	2	1	1
Black American	1	0	1
Latino	1	0	1
Native American	1	0	1
Unidentified	1	1	0
Average Age	23.2		23.2
Total	42	14	28

Table 1: Study 1 demographics, recruitment, dropouts, and total number (n)

B APPENDIX: STUDY 2 PARTICIPANTS

Table 2 presents information on Study 2 participants, encompassing demographics, recruitment, dropouts, and the total number.

Category	Recruitment	Dropout	n
Gender			
Female	20	5	15
Male	8	1	7
Race			
Asian	15	5	10
Caucasian	8	1	7
Multi-racial	1	0	1
Black American	2	0	2
Latino	2	0	2
Average Age	20.4		20.1
Total	28	6	22

Table 2: Study 2 demographics, recruitment, dropouts, and total number (n)

C APPENDIX: POST-TEST QUESTIONNAIRE RESULTS

Table 3 presents the mean values from the 7-point Likert scale responses in the post-test questionnaire.

D APPENDIX: SURVEY QUESTIONS

D.1 User Data

(1) How old are you?
(2) Do you have experience in teaching?
(3) If yes, how many years?
(4) How would you rate your familiarity with virtual reality?
  (a) 1: Not Familiar
  (b) 2
  (c) 3: Neutral
  (d) 4
  (e) 5: Familiar

D.2 Teaching Simulation Experience

(1) Which of the following conditions were you in?
  (a) Condition A: Normal Classroom (no gaze data visualization)
  (b) Condition B: Bar charts (gaze data visualized through bar charts)
  (c) Condition C: Fade-In (more opacity with more eye gaze)
  (d) Condition D: Fade-Out (more opacity with less eye gaze)
(2) Which of the following lectures did you teach?
  (a) Lecture A: Social Media's Impact on Society + Cybersecurity
  (b) Lecture B: Driving Manners and Road Safety + Public Transportation
  (c) Lecture C: Diversity + Sexual Misconduct
  (d) Lecture D: Alcohol and Other Drugs + Health and Wellness
(3) Mental demand: How much mental and perceptual activity was required to perform the task (e.g., thinking, looking, searching, deciding, teaching, etc.) (Scale from 1 to 100)?
(4) Physical demand: How much physical activity was required to perform the task (e.g., turning, controlling, etc.) (Scale from 1 to 100)?
(5) Temporal demand: How much time pressure did you feel due to the rate or pace at which the tasks occurred (Scale from 1 to 100)?
(6) Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals (Scale from 1 to 100)?
(7) Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance (Scale from 1 to 100)?
(8) Frustration level: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task (Scale from 1 to 100)?

D.3 Post-test Questionnaire

(1) Which condition was most helpful in improving your gaze behavior?
  (a) Condition A: Normal Classroom (no gaze data visualization)
  (b) Condition B: Bar charts (gaze data visualized through bar charts)
  (c) Condition C: Fade-In (more opacity with more eye gaze)
  (d) Condition D: Fade-Out (more opacity with less eye gaze)
(2) Which condition was least helpful in improving your gaze behavior?
  (a) Condition A: Normal Classroom (no gaze data visualization)

Study	Gaze Data Visualization	Reflecting on Gaze Behavior	Understanding Data Visualization	Tool Usefulness	Enhancing Nonverbal Teaching Behavior
Study 1	5.73	6.03	6.26	5.7	5.83
Study 2	6.31	6.40	6.45	5.86	6.13

Table 3: Average Values on the 7-Point Likert Scale in Post-Test Questionnaire

(b) Condition B: Bar charts (gaze data visualized through bar charts)
(c) Condition C: Fade-In (more opacity with more eye gaze)
(d) Condition D: Fade-Out (more opacity with less eye gaze)

(3) The gaze data visualization is easy to understand.
(a) 1: Strongly disagree
(b) 2: Disagree
(c) 3: Somewhat disagree
(d) 4: Neutral
(e) 5: Somewhat agree
(f) 6: Agree
(g) 7: Strongly agree

(4) The tool is useful in reflecting on my gaze behavior when teaching students.
(a) 1: Strongly disagree
(b) 2: Disagree
(c) 3: Somewhat disagree
(d) 4: Neutral
(e) 5: Somewhat agree
(f) 6: Agree
(g) 7: Strongly agree

(5) It is quickly apparent how to understand the data visualization.
(a) 1: Strongly disagree
(b) 2: Disagree
(c) 3: Somewhat disagree
(d) 4: Neutral
(e) 5: Somewhat agree
(f) 6: Agree
(g) 7: Strongly agree

(6) I consider the tool extremely useful.
(a) 1: Strongly disagree
(b) 2: Disagree
(c) 3: Somewhat disagree
(d) 4: Neutral
(e) 5: Somewhat agree
(f) 6: Agree
(g) 7: Strongly agree

(7) With the help of this product I will enhance my nonverbal teaching behavior.
(a) 1: Strongly disagree
(b) 2: Disagree
(c) 3: Somewhat disagree
(d) 4: Neutral
(e) 5: Somewhat agree
(f) 6: Agree
(g) 7: Strongly agree

(8) Did you feel like you knew what the data visualization (e.g. bar charts, opacity) represented?
(9) What did you like the most about the experience?
(10) What did you like the least about the experience?
(11) What, if anything, surprised you about the experience?
(12) What, if anything, caused you frustration?
(13) How would you describe your overall experience?