Modifying Pedagogical Agent Spatial Guidance Sequences to Respond to Eye-Tracked Student Gaze in VR

Adil Khokhar
University of Louisiana at Lafayette
Lafayette, Louisiana, USA
axk9375@louisiana.edu

Christoph W. Borst
University of Louisiana at Lafayette
Lafayette, Louisiana, USA
cwborst@gmail.com

Figure 1: Left: A student looks where the teacher agent points at after the teacher agent responds to their gaze drifting by asking them to look a little higher (a gaze trail is shown near the center of this image but is not visible to the student). Right: Teacher agent pauses and waits when pointing to blue barrels if student looks away.

ABSTRACT

In a guided virtual field trip, students often need to pay attention to the correct objects in a 3D scene. Distractions or misunderstandings of a virtual agent's spatial guidance may cause students to miss critical information. We present a generalizable virtual reality (VR) avatar animation architecture that is responsive to a viewer's eye gaze and we evaluate the rated effectiveness (e.g., naturalness) of enabled agent responses. Our novel annotation-driven sequencing system modifies the playing, seeking, rewinding, and pausing of teacher recordings to create appropriate teacher avatar behavior based on a viewer's eye-tracked visual attention. Annotations are contextual metadata that modify sequencing behavior during critical time points and can be adjusted in a timeline editor. We demonstrate the success of our architecture with a study that compares 3 different teacher agent behavioral responses when pointing to and explaining objects on a virtual oil rig while an in-game mobile device provides an experiment control mechanism for 2 levels of distractions. Results suggest that users consider teacher agent behaviors with increased interactivity to be more appropriate, more natural, and less strange than default agent behaviors, implying that more elaborate agent behaviors can improve a student's educational VR experience. Results also provide insights into how or why a minimal response (*Pause*) and a more dynamic response (*Respond*) are perceived differently.

© 2022 Association for Computing Machinery. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive version of record was published in ACM Symposium on Spatial User Interfaces (ACM SUI '22), https://doi.org/10.1145/3565970.3567697.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Virtual reality**; **Interactive systems and tools**.

KEYWORDS

virtual reality, gaze movements, 3D hotspots, pedagogical agents

ACM Reference Format:
Adil Khokhar and Christoph W. Borst. 2022. Modifying Pedagogical Agent Spatial Guidance Sequences to Respond to Eye-Tracked Student Gaze in VR. In *Symposium on Spatial User Interaction (SUI '22), December 1–2, 2022, Online, CA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3565970.3567697

1 INTRODUCTION

A guided virtual field trip can require that students are paying attention to the correct 3D locations of objects. Accurate 3D pointing is considered to be a key mechanism of effective communication and spatial guidance [3]; however, distractions or misunderstandings

[42] may cause students to miss critical information. For example, a student's attention drifts and they do not look when a teacher points out and explains a turbine. VR might address this problem somewhat with simple techniques such as visual cues to restore attention [43]; however, it's generally beyond the scope of current educational technology to adjust content based on visual distraction. To address this, we designed and evaluated an approach to make a pedagogical agent in VR responsive to student gaze. Our key contributions are:

- We present a VR pedagogical agent architecture that is responsive to a viewer's eye gaze based on eye-tracked visual attention during critical periods of time. An annotation system controls how prerecorded content can be sequenced in a timeline editor interface. The sequencing system then modifies playback of prerecorded content based on annotations to sequence more interactive (or appropriate) teacher avatar behavior.
- We demonstrate the success of our architecture in a virtual oil rig tour where a teacher agent points out and explains devices (see Figure 1) with an experiment that evaluated three different teacher avatar behavioral responses to student distraction based on eye-tracked visual attention. Student experience of a teacher avatar behavioral response with increased interactivity is compared to conventional teacher agent responses that consider minimal or no interactivity.
- We present an evaluation on the appropriateness of these teacher agent behavioral responses, along with insight into the importance of interactive agent behavior in an educational VR setting. Based on our findings, we offer guidelines for designing agent behavior and we provide insights into how or why a minimal response (*Pause*) and a more dynamic response (*Respond*) are perceived differently.

2 RELATED WORKS

2.1 Pedagogy in Education

Learners can miss out on critical bits of knowledge when distracted [30, 36] and continued distraction without any intervention may cause them to become disengaged or bored, leading to negligible learning gains from educational activities [15, 24, 29]. This lack of pedagogy can lead to a shallow understanding of educational material [16]. Studies done on pedagogy in education have included 2D educational video lessons and how the interactivity of controlling video playback can complement learning experiences for demonstrations or case studies [18, 21, 22]. Another study shows how dynamic video control and playback can complement learning activities and improve student understanding [44]. Another study investigated users' learning strategies with video playback activities such as: Selectively seeking in video clips to search and watch relevant content, pausing playback to attend to another activity such as writing a note or reflecting on what they have heard, and replaying a video clip to clarify something that they may not have understood [35]. The study found that during exam week, students would pause less and seek out segments with critical bits of knowledge in order to re-watch them. A study used 2D eye tracking to annotate critical note-taking content in lecture videos and then modified video playback by slowing or pausing when eye-tracking detected an activity such as note-taking [28] and found that students felt less cognitive load with gaze-reactive video playback.

2.2 Eye Tracking

Related works in this field focus on quantifying attention based on gaze patterns [31]. For example, one relevant 2D study establish links between attention and gaze movements by presenting distractors to draw gaze away from an object and measure reaction time as the duration it takes for gaze to return [33]. Recent studies have shown increased interest in eye tracking for understanding student attention in educational VR, such as investigating the effects of VR classroom configurations on students' attention [5, 19]. Eye tracking can provide mechanisms for a system to monitor and respond to shifts in attention [17]; however, less work has been done on assessing intervention strategies for a virtual agent's 3D spatial guidance in educational VR and whether gaze-reactive agents are more valuable than linear agents and/or simple system interruptions. Similar related work focuses on 2D gaze-sensitive dialogs used in desktop interfaces to redirect students' attention to important areas and show that gaze-reactivity is effective in promoting learning gains for deep reasoning questions. Our work follows similar methodology: Studies typically will first assess how different approaches to sequencing agent guidance are perceived, and then assess attention and learning in a later study after the basics are understood [12, 13].

2.3 Educational VR

Immersive 360-degree videos have shown promise by producing better learning outcomes compared to traditional video instruction by being more interactive and allowing students to easily identify critical information in less time [9, 40]. Recent works have explored using instructors in virtual educational experiences to tutor machine operating tasks [8] and demonstrate laboratory procedures [37]. Similar works have developed educational VR experiences that taught 3D design tasks with a video tutorial system [39], used an AR visualization system to assist an instructor teaching students in a VR classroom [38], facilitated educational lecturing and collaboration with a VR whiteboard tool [20], and had high school students take a virtual field trip to a virtual 3D solar energy center to teach concepts about solar energy generation [4]. This is representative of virtual field trips of large environments or structures, an application for which VR can be well-suited. Virtual oil rigs have been used previously to train workers [34] and assess the impact of distractions on attention [43]. Therefore we consider a virtual oil rig environment where a teacher agent points out and explains equipment that is used on oil rigs.

3 METHODS AND SEQUENCING ARCHITECTURE

We extended a preliminary version of an avatar sequencing framework presented by Khokhar et al. [23]. The sequencing system processes response descriptions and executes the teacher agent response. The main idea behind our sequencing system is like a subsumption architecture from robotics [6]. Behaviors are selected using sensor-based conditions and behavior priority. Unlike the

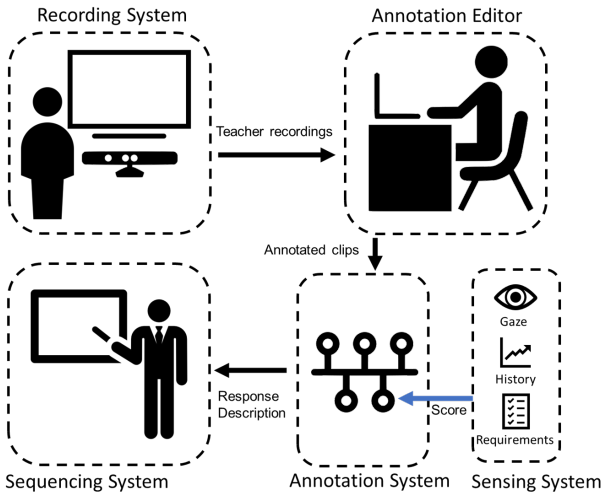

Figure 2: How our architecture sequences teacher clips into teacher responses. First the teacher records some clips. The clips are annotated by an annotator. The sequencing system interprets the response description composed by the annotation system and generalized hotspots to sequence appropriate teacher avatar behavior and system responses.

Figure 3: The teacher agent points to a deck crane and waits for the student to look. A gaze trail (not visible to student) shows student attention shifting from teacher to the pointed-at object, thus resuming default teacher behavior.

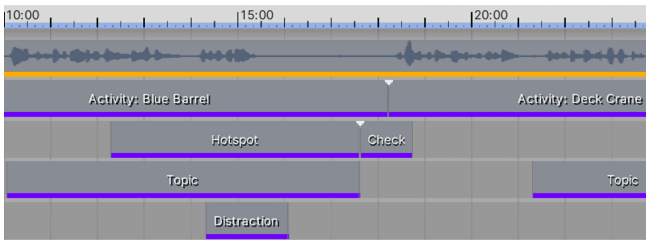

Figure 4: Associated annotations on our editor integrated into Unity's Timeline affecting sequencing of teacher avatar clip playback to create teacher agent behavior.

original architecture, we do not use weighted random determination of responses. Triggering agent behavior based off changes in sensor data alone is not likely to be helpful. Furthermore, we want the agent to react only when it is appropriate to do so. These considerations led us to create an architecture that dynamically modifies the sequencing and playback of prerecorded clips to create appropriate teacher agent behavior (see Figure 2). The *sensing system* allows the agent to sense attentional shifts by the use of generalized hotspots. The *recording system* handles the creation of prerecorded teacher agent content. Our *annotation system* produces a response description that contains candidate teacher avatar responses by composing a combined distraction score with annotations. The *sequencing system* is the simplest layer, it receives the response description, computes the final rank, and sequences the response. As a simple example, in Figure 3 a teacher points at a deck crane and a gaze angle hotspot with associated annotations sets up pausing of teacher playback until the student looks. If the student's gaze continued to drift, a history hotspot adds to the score to trigger the sequencing of an alternative clip where the teacher tells the student where to look based on the hotspot's sensed gaze direction.

3.1 Recording System

Our teacher agent has a 3d RGBD-based avatar built by prerecorded color and depth videos captured by a Kinect V2 at 30 FPS. The implementation for our FFmpeg-based Unity plug-in is similar to the implementation from Ekong et al. [14], but we also capture movement/orientation pose frames of the Kinect V2 skeleton that are then mapped to a Unity Mecanim Humanoid Rig for purposes of identifying critical periods of time when the teacher is pointing.

Content in recordings was produced such that the main presentation topic is first and alternate clip segments that compose responses came at the end. Each clip has 6 different sub-clips: 4 where the teacher asks the student to look towards each particular direction (left. right, up, down), 1 where the teacher asks the student to look where the teacher is pointing, and 1 to tell the student that he will go over the explanation again.

3.2 Sensing System

Our system generalizes beyond standard response-triggering hotspots (e.g., circular gaze targets in Google Expeditions) in order for our agent to express more complex behavior than just a pre-recorded clip being triggered or paused by a gazing at a hotspot. Whereas standard hotspots directly trigger a response, the implementation of our generalized hotspot composes multiple low-level sensors to check more complex conditions than traditional hotspots. For example: an "inverse" hotspot where everything beyond some angle to a user is the hotspot. Additionally our hotspots can consider 3D spatial locations, generalized hotspots can be associated with multiple scene areas (objects), and our hotspots provide a mechanism to define requirements to progress an educational activity. Our

teacher agent senses attentional shifts through generalized hotspots built from the following components that combine low-level sensor information: Sensors, Mappers, and Combiners.

Sensors measure low-level properties or device information including device data, for example, pupil diameter from an eye tracker API. Typically, each sensor receives a single input and passes it to a combiner system. Other sensor types are included to support specific sequences in educational activities, e.g., controller input and aim, agent pose, the state of a game object, and angle of the user's gaze away from an object or teacher avatar (by computing $\theta = 2 \cdot atan2(||u - v||, ||u + v||)$ [10] where u and v are two relevant direction vectors).

Mappers transform the input by a function, such as a temporal filter, a low-pass filter, or a nonlinear response curve, e.g., a mapper applies temporal smoothing to the eye gaze angle using a first-order IIR filter, with a filtering rate of 60Hz (Unity update rate), to remove effects of very brief eye movements or jitter: $y_n = 0.9 \cdot y_{n-1} + 0.1 \cdot x_n$ [1]. Also, in our study, we map gaze angles so that 25° and 90° map to minimum and maximum scores.

Combiners receive inputs and apply operators that transform multiple inputs into a single output, e.g., a combiner receives two angles describing the user's gaze deviations from the teacher avatar and from the pointed object. The combiner computes the minimum of the two angles so a user can be considered attentive if looking at either.

By default, a generalized hotspot computes an distraction score using angle difference sensors. Combiners allow for multiple hotspots and combine multiple scores that check conditions for outstanding prompts, activity repetitions, incorrect/correct quiz questions, and references to other hotspots and their requirements. The methods follow those of Khokhar et. al [23]. In Figure 3, a teacher agent points at a deck crane and a generalized hotspot with a gaze angle sensor has been annotated to set up pausing of teacher playback until the student looks. If the student's gaze continued to drift, a history sensor eventually sufficiently contributes to the score to trigger the sequencing of an alternative clip wherein the teacher tells the student where to look based on the hotspot's gaze direction sensor. The activation and control of these generalized hotspots are based on annotations.

3.3 Annotation System

The annotation system is responsible for computing a response description for the sequencing system based on the combined distraction score and annotations. Annotations (metadata) can represent timeline annotations, specify timing of responses, timing of clip content, critical periods, activation and control of generalized hotspots, player history, timers, candidate responses, and agent execution history to compose a response description that contains a combined distraction score and candidate teacher avatar responses. Unity's Timeline feature is used to coordinate audio, animation, and game object activations. We extended this to support an editor for annotating teacher content with metadata. Associated annotations on our editor are integrated into Unity's Timeline and affect sequencing of teacher avatar clip playback to create teacher agent behavior. The annotation system processes annotation tracks and active hotspots to generate a response description with a combined distraction score and affects candidate responses based on active annotations for the current point in time. This response description is then ranked by the sequencing system. Annotation tracks differ from standard timeline tracks in that they provide contextual information to the response description and allow an editor to adjust how prerecorded content on the timeline is sequenced, such as arranging or modifying the playback, seeking, rewinding, and stopping of teacher clips in Unity's timeline editor. Five types of annotations control the sequencing and playback.

A Hotspot annotation is a critical period of time, specified in clip timestamps, that controls the activation and timing of a generalized hotspot. By default a generalized hotspot has a 500ms activation period. This can be configured in the editor.

An Activity annotation is a clip-level property that associates prerecorded content with a group of annotations, generalized hotspots, and can represent a candidate response for the agent. For example, hotspots can determine which way a user needs to turn to look at an object based on the user's current gaze. The activity annotation composes the teacher agent response that asks that student to look towards a specific direction. If the gaze angle history hotspot indicates continued distraction or it is ambiguous which direction to ask the student to look to, then the teacher agent will seek to the activity annotation that represents a response asking the student just to look where he is pointing.

A Topic annotation is an interval of time, typically a sentence, that allows referencing specific times in a candidate response by providing timing information. For example, suppose the student's gaze continues to wander as the teacher points to a deck crane, requiring the currently executing response to be interrupted but only when it is next appropriate to do so. The topic annotation of the current response is compared to the target response to determine when is most appropriate to seek, e.g., at the end of the sentence after pointing, and where to seek to, e.g., at the beginning of a sentence.

A Check annotation is a critical point in time where teacher playback behavior can be modified to create a teacher avatar behavioral response if the hotspot conditions were not met. For example, the teacher points at a blue barrel and the hotspot requires that the student's gaze angle to the pointed object is below a certain threshold. The student does not look at the pointed at object so the sequencing system modifies teacher avatar clip playback behavior so that a response takes higher priority than the default response of continued playback.

A Promote Responses annotation can be configured as a clip-level property or a critical period of time. This promotes any particular activity to candidate response and affects ranking in the response selection mechanism. Omitting the period of time makes hotspots in the activity always-on. For example: an always-on hotspot detects the student's gaze away from the current educational object for displaying an attention-guiding arrow. Another example is that when the teacher asks the student to answer questions on their mobile device, this annotation promotes a response that pauses playback and subsumes all other candidate responses until the student has met all requirements for the active hotspot by completing all required tasks and answering all quiz questions.

3.4 Sequencing System

The sequencing system receives the response description and uses the techniques described earlier to calculate a rank and choose the highest-utility response, then modifies the playing, seeking, rewinding, stopping, and pausing of a teacher recording accordingly.

For example, suppose the agent points out a specific button on a handheld controller, to be pressed by the student as practice. The timeline includes the Hotspot, Topic, Promote Responses, and Check annotations. A hotspot senses if the student has not pressed the button. A Promote Responses annotation promotes certain candidate responses by elevating their ranks and specifies that currently executing behavior should be subsumed if the requirement is not met when the teacher points at the controller and the Check annotation is reached. Suppose, optionally, the responses are to be constrained in order: pause, replay, or play a different clip. The constraint is applied using response ranks. The teacher first pauses when pointing and a cooldown timer prevents other responses from activating by temporarily demoting ranks so the student has time to meet the requirement. After the timer ends, the execution history demotes the pause's rank. Then the next response "Replay" can be selected and another cooldown timer starts. Replay causes the teacher repeat the sentence and point once again. If the student again doesn't meet the requirement, the execution history demotes replay and the next response will make the teacher play a different clip acknowledging the student's inability to complete the task and perhaps offering help or moving on to a different activity. The combined distraction score affects computation of ranks. If the score is very high, e.g., a very distant gaze or a history of extraneous inputs, then the last response from the order above, playing a different clip, can be promoted in rank to subsume other responses immediately. With increasingly complex examples, the benefit of this AI architecture is that the agent can be more dynamic and extensible without an extensive set of explicit if-then conditions.

3.5 Teacher Avatar Responses

Our three teacher agent behavioral responses were inspired by real-world communication strategies for distraction: *Continue*, *Pause*, and *Respond*. *Pause* is more representative of simple interactions in research that has investigated minimally interactive interventions [19] and is modeled off of maintaining eye contact. *Respond* is inspired as a variation of an Initiate Respond Evaluate (IRE) sequence where an instructor asks a question, the student responds, and the tutor evaluates and provides feedback to the response [27]. A minor difference from this is that our teacher agent offers a reminder of the main task by repeating the instruction after sufficient additional direct instruction has been provided. These teacher avatar behavioral responses drew inspiration from students' learning strategies during exam week with educational video playback such as manual pause and rewind controls [28]; however, automating them could be more seamless and ensure a response for students that either are unable to locate segments with critical knowledge or forget to pause or rewind.

We composed 3 teacher avatar behavioral responses using our architecture: *Continue*, *Pause*, and *Respond*.

The *Continue* response is modelled after a linear agent. The teacher avatar continues to play through the recording. For example, the teacher avatar points at the iron roughneck, the student may or may not look at the iron roughneck, and the teacher avatar continues to explain what the iron roughneck does.

The *Pause* response is modelled after a reactive agent with a minimal level of interactivity. The teacher avatar pauses and waits for the student to be ready to continue. For example, when the teacher avatar asks the student to answer questions on their mobile device, an annotation activates a generalized hotspot that affects sequencing so a teacher agent's *Pause* response subsumes all other behaviors until the student has completed all required tasks and answered all quiz questions.

The *Respond* response is modelled after a reactive agent with a higher level of interactivity. The teacher pauses for a second, then responds to the student. For example, the teacher explains the drill string and points at it. If the student does not look at the drill string, the teacher will ask the student to look a little bit to the left (for example). Continued distraction will cause the teacher to rewind and repeat instructions; however, if attention is regained then the teacher will skip to the next topic if appropriate.

We provide an example of how the annotation system works with the *Respond* condition in Figures 3 and 4. When the Check annotation is encountered in the timeline and the *Respond* behavioral response condition is active, a generalized hotspot is activated by an annotation and compares directional components between gaze direction and the pointed object. If the student needs to only turn slightly towards the object, then an appropriate clip is sequenced to play where the teacher notifies the student to look in that direction. Otherwise the teacher will ask the student just to look where he is pointing. Finally if the student continues to be distracted, the teacher will say "Let's try going over that again" and reminds the student to look at where he is pointing, before rewinding to the start of the topic. If the student's gaze continued to wander during a critical period what will happen depends on the teacher behavioral response condition: During *Continue*, the teacher will continue playing. During *Pause*, the teacher will pause and wait on student until they resume looking. During *Respond*, the teacher asks the student to look at the object they're pointing at, continued distraction will then have the teacher provide a hint as to where to look, and finally if gaze continues to wander then the teacher will tell the student that he's going to go over it again and repeats the instruction.

3.6 Virtual Mobile Device

To support our experiment, the observer holds a virtual mobile device in our software for general use, such as interactions, and as a mechanism for the system to present distractions in a controlled manner to support experiments. The virtual mobile device provides a way to experimentally simulate distractions, without easily being ignored, and with a standard interaction tool that fits the theme.

3.6.1 Distractions. Our virtual mobile device presents distractions by showing them as text messages with accompanying vibration and sound effect. A low level distraction is a text message that does not require a response. A high level distraction is a text message that requires interaction with an object, such as pointing at an object. Figure 5 illustrates our two levels of distractions.

Figure 5: Left: A low level distraction that does not need to be acknowledged. Right: A high level distraction requiring the student to interact with a deck crane.

For our experiment evaluating teacher responses, distractions are timed to occur shortly before critical periods in a randomized manner, such as when the teacher points at an object they're explaining. The timing of these distractions and modifying of teacher behavior when a student does not acknowledge the high level distraction are handled through annotations (Section 3.3).

4 EXPERIMENT DESIGN

We evaluated our architecture with a within-subjects user study that assessed and compared the subjective suitability of teacher agent behavioral responses. The main independent variable was the type of teacher agent behavioral response (*Continue*, *Pause*, and *Respond*, see Section 3.5). Subjects consisted of 31 male and 6 female students that were recruited from our Computer Science department for a total of 37 subjects: 28 undergraduates and 9 graduate students. Ages ranged between 18 to 40 years with a median of 21. 9 subjects indicated prior experience with VR field trips. 10 out of the 37 subjects indicated that they owned a VR headset. In addition, 1 subject indicated prior knowledge of devices used on the oil rig. The apparatus for the experiment included a Vive Pro Eye headset with a facial tracker attached, a Vive wand, a logitech R400 clicker, a large Samsung TV, and a desktop with an Intel Core i9 10900K CPU processor, GeForce 2080 graphics card, and 64GB of memory. Our study consisted of three phases of watching a teacher give an educational presentation.

4.1 Procedure

4.1.1 Setup and Calibration. After signing a consent form, the subject was given an experiment overview of three phases before donning a Vive Pro Eye headset. The proctor checked that subjects reported being comfortable and seeing clearly before starting a calibration. During eye tracker calibration, the subject adjusted their headset until the eye tracking sensors detected their eyes, adjusted the inter-pupillary distance dial to the appropriate setting, and looked between 5 points in their visual field. The system compared gaze measurements to those of the predetermined points and adjusted device configuration based on the deviation. After calibrating, the subject looked at spheres that were arranged in a grid and the proctor confirmed accuracy of detected eye gaze by looking at a particle trail that followed the subject's eye gaze ray. Subsequently, the particle trail was disabled and the student was given a tutorial by the in-game teacher agent that introduced them to the oil rig and in-game controls.

4.1.2 Phase 1: Presentation Phase. In Phase 1, the subject was given an educational presentation that visited six oil rig areas, guided by the teacher agent. Teacher agent behavior conditions were randomly ordered according to a latin squares design of size three. The six oil rig areas were grouped into two sets of three consecutive areas per set. Within each set, each area presented a different agent behavior. In each area, the teacher avatar pointed out and explained three oil rig devices. The virtual mobile device presented a distraction two of the three times, with one high level and one low level distraction. The distractions were randomly ordered so each teacher behavior condition was paired with a low-high or a high-low distraction order across the two sets of areas for a total of six teacher behavior-distraction level order conditions. This randomization of distractions was to reduce subject anticipation of a distraction every time the teacher avatar pointed and asked the subject to look at an object. The subject was required to gaze at the correct object when the teacher agent avatar pointed out a device. At the end of the area, if the subject did not acknowledge the high level distraction, then the teacher avatar reminded the subject to address the distraction and waited until the subject acknowledged the distraction. The subject then answered two quiz questions about educational content presented in the area and rated the quality of the teacher avatar behavior. The last prompt asked the subject to give feedback on any strange or unusual teacher behavior. The proctor then conducted an open-ended interview to obtain any additional feedback on the subject's ratings. To ensure that the subject provided feedback, the proctor was required to press a key before the subject could continue to the next area. At the end of Phase 1 the subject took off the headset and was provided a break.

4.1.3 Pre-Recorded Student. Playback of prerecorded student and teacher sessions were used in later phases, showing the subject a distracted student to observe in order to evaluate appropriateness of teacher agent behavior to two levels of prerecorded student distraction length: low duration and high duration. The subject reviewed six recordings of prior sessions of a distracted student in the oil rig while different teacher responses occurred. The six recordings were divided into two sets, with three recordings per set. The student recording captured all movement/orientation data and annotations about the state of the game environment. In recordings chosen for review, the recorded student appeared distracted for either a brief moment (low duration) or long enough that a teacher agent in the *Respond* condition offered guidance and then repeated their instructions (high duration). We list student distractions below:

(1) Condition Set 1:
   - Student stared at teacher and is unresponsive for a brief moment.
   - Student looked down at phone for a brief moment.

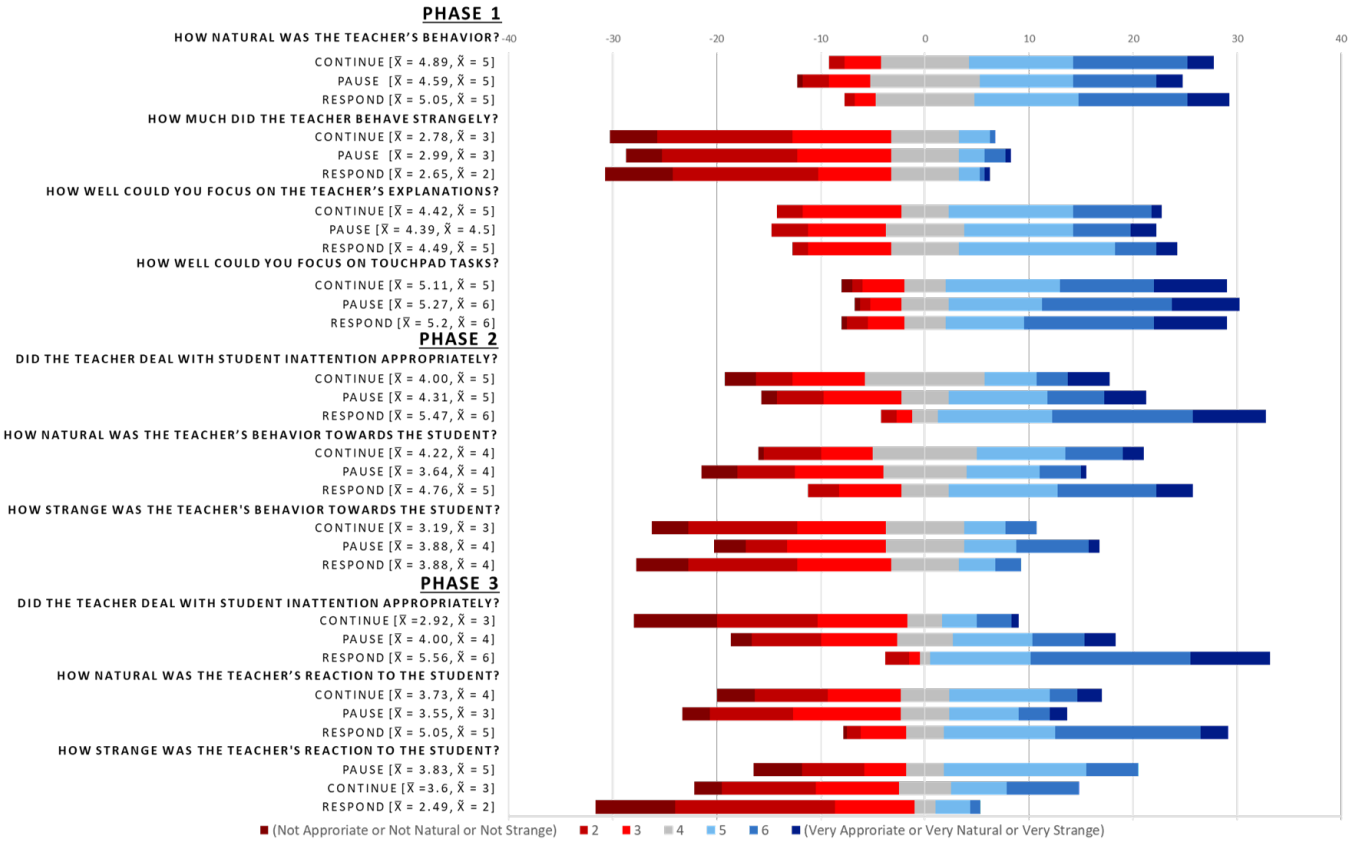

Figure 6: Stacked Bar Charts, Counting Responses to Teacher Rating Questions. Each bar has a width of 37, the number of total subjects, and consists of colored segments that show how the responses were distributed. Segments are arranged such that neutral responses are centered at 0 so that the bar's horizontal range relates to the number of positive (right of 0) and negative (left of 0) responses.

- Student pointed phone wand towards rig object and did a point and click task. This was a longer distraction.

(2) Condition Set 2:
- Student looked at wrong rig object for a brief moment.
- Student moved phone up and looked at it for a brief moment.
- Student pointed phone wand towards rig object and did a point and click task. This was a longer distraction.

4.1.4 Phase 2: Observer Phase. In Phase 2, the subject was placed in the role of an observer that reviewed a recorded student's experience from the recorded student's first-person perspective. The purpose was to assess different possible teacher responses in a more controlled manner (not hinging on the subject's behavior). To avoid discomfort associated with seeing another person's first-person view, the subject viewed Phase 2 on a large television (75 inch Samsung), and 6-tap filtering of playback head orientation removed high-frequency jitter. Subjects selected ratings and confirmed prompts with a Logitech R400 input device. Phase 2 consisted of two sets with three areas per set. Student distraction in each area had either a low duration condition or a high duration condition. The prerecorded student first showed a low-distraction duration in two areas and then a high distraction duration condition in the third area, where the teacher response was longer. Each set consisted of the 3 different teacher conditions with a randomized order based on latin squares. The teacher recordings were abbreviated versions of clips from Phase 1. The teacher clip and the pre-recorded student clip were played together as they occurred together when the student was recorded. At the end of each visited area, the subject rated the teacher avatar's responses. At the end of every set of 3 clips, a prompt appeared and the proctor asked the subject to rank the 3 teacher behaviors of the 3 clips they watched.

4.1.5 Phase 3: Reviewer Phase. In the reviewer phase, the subject returned to headset-based VR and reviewed clips of the different teacher avatar responses to a pre-recorded student experiencing a distraction (three possible teacher responses to the same pre-recorded student clip) of short duration. This focused the subject on teacher response options during a single student distraction. In this phase, the subject was an external observer in VR with an "over-the-shoulder" view behind the (pre-recorded) student, with both the student and teacher in view. The pose of the observer relative to the student was the same through each of 3 areas. For each of the 3 areas, the teacher response condition was shuffled

Table 1: Subscale scores (using averages of contributing questions rating teacher response from Figure 6) and statistical comparison of *Continue* response to *Pause* response, *Continue* response to *Respond* response, and *Pause* response to *Respond* response with Friedman tests followed by post-hoc Wilcoxon Signed-Ranks Tests.

		Friedman		C vs. P		C vs. R		P vs. R	
	Averaged Subscales	$\chi^2(2)$	**p**	**Z**	**p**	**Z**	**p**	**Z**	**p**
Phase 1	Natural	8.407	**.015***	-1.705	.088	-1.646	0.1	-3.3	**.001***
	Strange	6.323	**.042***	-1.211	0.226	-1.041	0.298	-2.155	**.031****
	Teacher	0.391	0.823	-0.269	0.788	-0.499	0.617	-0.285	0.775
	Touchpad	2.431	0.297	-1.041	0.298	-0.909	0.363	-0.41	0.682
Phase 2	Appropriate	22.141	**<.001***	-1.571	.116	-3.880	**<.001***	-3.834	**<.001***
	Natural	25.526	**<.001***	-2.513	**.012***	-2.448	**.014***	-3.947	**<.001***
	Strange	13.318	**.001***	-2.812	**.005***	-.905	.365	-3.549	**<.001***
	Rank	35.504	**<.001***	-.279	.780	-4.823	**<.001***	-4.317	**<.001***
Phase 3	Appropriate	39.986	**<.001***	-2.832	**.005***	-5.186	**<.001***	-4.122	**<.001***
	Natural	25.051	**<.001***	-.637	.524	-3.854	**<.001***	-4.256	**<.001***
	Strange	20.561	**<.001***	-.728	.467	-3.412	**.001***	-4.058	**<.001***
	Rank	39.135	**<.001***	-1.945	.052	-5.090	**<.001***	-4.407	**<.001***

** Items displaying a significant difference are marked in bold and followed by an asterisk.*
*** This value would not survive Bonferroni correction but is accepted as the only significant followup to the Friedman result*

into a random order and the subject reviewed a clip of the teacher responding to a student, for a total of 9 trials. The prerecorded student was distracted by a point and click task and the teacher reacted according to the condition. At the end of each teacher clip, the subject gave their ratings. Then the teacher clip cycled to the next teacher response condition and played back the same student event. At the end of an area, when all 3 teacher response conditions were shown, the proctor asked the subject to rank the responses.

5 RESULTS AND DISCUSSION

5.1 Ratings

We compared the subjective suitability of the 3 teacher agent response conditions assessed based on responses given to questions listed in Figure 6. Except where stated otherwise, question responses were ratings from 1 to 7. Semantic anchors were placed below values 1 and 7, with 4 being a neutral answer. We found several significant effects in phases 2 and 3 when comparing response conditions. We present the responses to both standard questionnaire items and to in-game scores as stacked bar charts that count the various responses summarized in Figure 6. Each bar has a width of 37, the number of total subjects, and consists of colored segments that show how the responses were distributed (see the figure caption further describing the plot layout). Ratings were first analyzed with Friedman tests followed-up up by Wilcoxon signed-rank tests. We split up the summary into a list addressing each rating metric individually with Table 1 summarizing results.

5.1.1 Phase 1: Presentation Phase. Phase 1 ratings are summarized in Figure 7 per question and teacher response condition and in Table 1. Ratings of *naturalness* show differences between teacher response conditions based on a Friedman test ($\chi^2(2, 74) = 8.407$, p=.015). Post-hoc tests reveal that *Pause* was found to be less natural than the *Respond* condition. When asked what contributed to low natural ratings for *Pause*, 9 subjects felt that the teacher's pausing animation while pointing felt unnatural, with 2 subjects further adding that the duration of *Pause* was too long, suggesting a need to improve the interactivity of *Pause*, such as by playing an idle animation. Ratings of *strangeness* differ ($\chi^2(2, 74) = 6.323$, p=.042) between teacher response conditions. Post-hoc tests show that subjects found *Pause* to be stranger than *Respond* or *Continue*. Note that this is the only comparison that would not survive Bonferroni correction but for consistency with the Friedman result we accept it as significant. When asked if they noticed anything strange or unnatural, 17 subjects stated that they did not notice anything

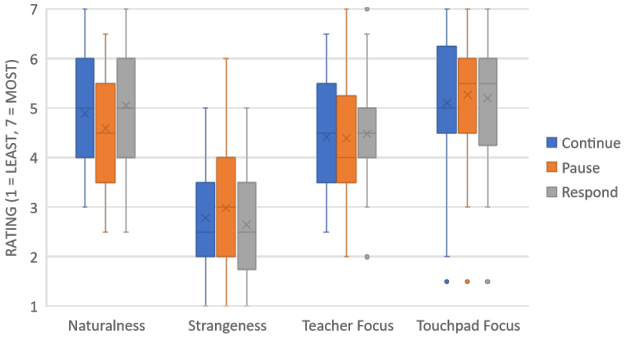

Figure 7: Phase 1 - Scores for different teacher behaviors. Box-and-whiskers plot showing median, interquartile boundaries, and outliers (asterisks).

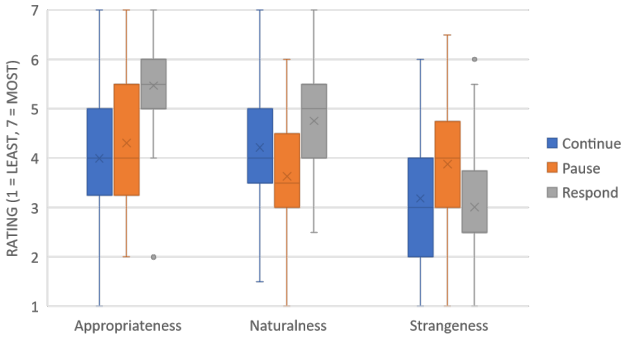

Figure 8: Phase 2 - Scores for different teacher behaviors.

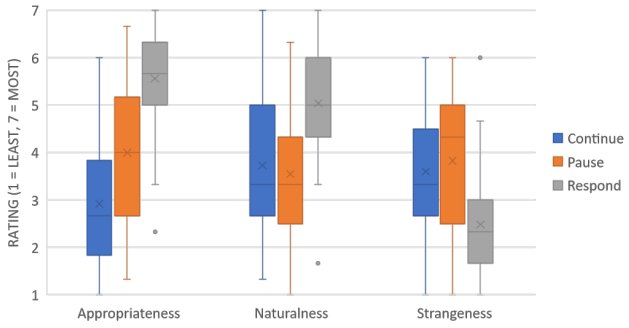

Figure 9: Phase 3 - Scores for each teacher behavior.

strange or unnatural for all 6 areas. Open-ended interviews with subjects revealed 10 subjects that said they could not focus on the teacher due to the phone distractions. Answers to quiz questions were encoded as correct or incorrect based on the subject's answer. A related-sample Cochran's Q Test did not show any significant differences between teacher response conditions ($\chi^2(2, 74) = 4.667$, p=.097) or distraction level order ($\chi^2(1, 74) = .286$, p=.593). The purpose of quizzes was to encourage student attention and we did not expect significant differences.

Although not detailed in this paper, we additionally did not find any significant differences in any of the other rating items. This lack of significant differences between *Continue* and other teacher responses might be related to subject interviews revealing that it was not too uncommon of an experience for a real human teacher to ignore student inattention and continue on with a lecture. We additionally checked some eye gaze metrics per condition between subjects: number of blinks per minutes, angle between the direction of the subject's gaze direction and the direction from the subject's eye center to the teacher avatar's head, and time spent looking at the teacher and pointed-at objects. We did not detect any significant differences when comparing summary values (mean, median, and standard deviation) because these metrics had high variance across subjects.

5.1.2 Phase 2: Observer Phase. For phase 2 we summarize the results shown in Figure 8 and Table 1. Ratings of *appropriateness* differ ($\chi^2(2, 74) = 22.141$, $p < .001$) between response conditions with post-hoc tests showing *Respond* ranking higher than both *Continue* and Pause. Ratings of *naturalness* differ ($\chi^2(2, 74) = 25.526$, $p < .001$) between response conditions with post-hoc tests showing *Pause* being the least natural and *Respond* being the most natural. Ratings of *strangeness* differ ($\chi^2(2, 74) = 13.318$, $p = .001$) between response conditions with post-hoc tests showing *Continue* to be stranger than *Respond*. When asked what contributed to high *Respond* ratings, subjects felt that the teacher's responsiveness to student inattention guided attention back appropriately. Subjects liked how the teacher would guide back attention by reminding the student where to look. When the student was distracted for longer, subjects also liked that the teacher avatar would repeat himself if the student's inattention continued. This is in contrast to 25 subjects who felt the long duration and stillness of *Pause* made it especially noticeable and awkward. This is reinforced by 16 subjects who felt that *Pause* did not do anything to address student inattention. Finally, 20 subjects felt that *Continue* was awkward because it did not address student inattention. Additional post-hoc Wilcoxon Signed Rank tests revealed a significant difference between the *Pause* and *Continue* responses particularly during high distraction (Z = -3.877, p <.001); however, the same difference is not detected with low distraction (Z = -.103, p = 0.918). This suggests that subjects felt long pauses to be especially noticeable and unnatural. Open-ended discussion with subjects revealed that a real human teacher would not remain paused and eventually continue. This suggests that pausing behavior could be improved with an idle animation or other behavior.

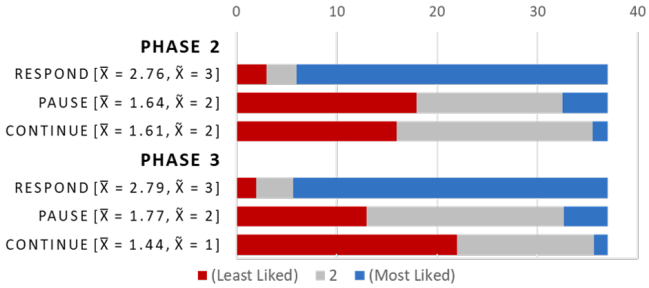

Figure 10: Preference ranking results for the three teacher avatar behavioral responses in phases 2 and 3. Each bar consists of colored segments that shows the distribution.

5.1.3 Phase 3: Reviewer Phase. For phase 3 we summarize the results shown in Figure 9. We found several significant effects when comparing teacher avatar response conditions.

Ratings of *appropriateness* differ ($\chi^2(2, 111) = 39.986$, $p < .001$), with *Continue* ranking the lowest, *Pause* being in-between, and *Respond* ranking the highest. Ratings of *naturalness* differ ($\chi^2(2, 111) = 25.051$, $p < .001$), with *Pause* ranking the lowest, *Continue* being in-between, and *Respond* ranking the highest. Ratings of *strangeness* differ ($\chi^2(2, 111) = 20.561$, $p < .001$), with *Pause* ranking the lowest, *Continue* being in-between, and *Respond* ranking the highest. When asked what contributed to low ratings for *Pause* and Continue, subject interviews revealed that 11 subjects felt that the long duration and pointing frozen animation of *Pause* was awkward,

with 11 subjects stating that the unresponsiveness of *Pause* did not do anything to address student inattention. 23 subjects felt *Continue* did not address student inattention, causing the student to miss out on important information. *Respond* was found to be the most natural for reasons similar to those in phase 2. We found significant effects for reasons similar as to what was discussed in *naturalness*. Subjects felt it was strange for a teacher avatar not to respond to the student's inattention. Subjects had mixed preferences on the *appropriateness* of *Pause* compared to *Continue* as they felt that *Pause* was more appropriate for shorter durations of distraction and *Pause* was better because the teacher avatar reacted in some way rather than not at all.

5.2 Ranking Results

We assigned each teacher behavior condition an integer rank from 1 to 3 based upon subject least-liked and most-liked responses, with 1 being least liked and 3 being most liked. Figure 10 summarizes the results from all subjects. Rankings were averaged and analyzed with Friedman tests with Wilcoxon signed-rank followups. Results of analyzes are shown in Table 1. Phase 2 ($\chi^2(2, 74) = 35.504$, $p < .001$) rankings differed between response conditions, with post-hoc tests showing *Respond* ranked higher than *Continue* or *Pause*. Phase 3 ($\chi^2(2, 111) = 39.135$, $p < .001$) rankings differed between response conditions, with post-hoc tests showing *Respond* ranked higher than *Continue* or *Pause*. In phase 2, when asked what contributed to high *Respond* rankings, the majority of subjects (22) felt that *Respond* acknowledged student inattention and guided it back. This is reinforced by 15 subjects that felt *Respond* clearly told the student what to do and was very easy to understand. In phase 3, when asked what contributed to high *Respond* ratings, the majority of subjects (30) echoed sentiment from Phase 2 that *Respond* acknowledged student inattention and guided it back.

6 CONCLUSION AND FUTURE WORK

This work designed and evaluated an approach to make a VR pedagogical agent responsive to loss of a student's visual attention. Our annotation system provides an interface that associates teacher avatar responses with prerecorded content and controls its sequencing by subsuming default playback behavior based off of generalized hotspots to improve virtual teacher responsiveness.

We tested the approach with a virtual oil rig tour wherein an agent sequences prerecorded teacher clips to point out and explain devices. Results suggested that users considered teacher avatar behavior with increased interactivity to be more appropriate for addressing student distraction, more natural, and less strange, compared to behaviors that have minimal or no interactivity. Merely pausing a clip to wait for a student is less preferable and, in some cases, may even leave students more distracted or confused.

While it is not surprising that more elaborate agent responses are useful, such as *Respond*, we believe that including *Pause* was important for understanding tradeoffs between minimal and more dynamic behaviors. We note that a few subjects in our study disliked more interactive agent behavior (see Figure 10), with their feedback suggesting that it was too direct. There is some evidence that students with low prior knowledge or a high rate of errors may learn better with more polite and less direct instruction [25, 26]. Based on our findings, we offer some design guidelines for pedagogical agent developers that can improve user experiences: Agent behavior considering minimal interaction should, at least, indicate to a user that the teacher agent is aware of their distraction and that the agent's behavior has changed, for example, by fading out or playing an idle animation on the teacher avatar. Such behavior can be appropriate when a student restores attention after experiencing a brief interruption or the teacher agent is waiting on the student to complete a task. More elaborate teacher agent behavior, such as *Respond*, should be considered when direct intervention or guidance is required to address student distraction. For example, when a student misunderstands how a motor works and the teacher agent provides additional clarification, or when a student becomes inattentive from a distraction and the teacher redirects their attention back to the relevant object. Compared to default pedagogical agent behaviors, our suggested guidelines can lead to more natural and dynamic pedagogical agent behavior and may even detect student difficulties or misunderstandings that are not realized by the student.

Our work chose simple behaviors as they were a good starting point to test and understand tradeoffs, as well as gather insights. Future work can consider building more elaborate behaviors that model dialogues from expert tutors by extending our annotation system with the scheme presented in [7] and assess the effect on student learning. Other improvements for our teacher behaviors can be considered such as fading out the main teacher avatar and playing back a second teacher avatar for *Respond* to more overtly indicate to the student that the teacher's behavior has changed. Another aspect future studies could consider is improved user-controlled interactivity of *Pause* by allowing the subject to choose different levels of playback control, such as manually pausing, resuming from a play button, or even combining with attention restoration guiding cues [43]. Improving playback control may further prevent students from missing critical information during distractions.

Although our focus is on the student experience in this work, future work can consider techniques to extract teacher guidance and intent from recordings to automatically generate preliminary annotations, for example, by detecting pointing in the clips and setting up initial annotations. These would then be tuned or extended by a designer or teacher, but deployed to very many students afterwards. While our study did not detect significant differences in eye gaze metrics due to high variance across subjects (Section 5.1.1), future work can consider improved detection of student attention by incorporating more advanced consideration of eye behavior or even other sensors. Hotspots could be extended to include machine-learning-based detection [2], physiological sensing such as electrodermal activity [11, 41] or electroencephalography [32]. Finally, integrating gaze-responsive techniques with agents that use natural language generation and conventional avatars instead of prerecorded clips may extend results to other avatar types.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1815976.

REFERENCES

[1] J Arzi. 2016. Tutorial on a very simple yet useful filter : the first order IIR filter. (April 2016), 8. http://www.tsdconseil.fr/tutos/tuto-iir1-en.pdf

[2] Sarker Asish, Ekram Hossain, Arun Kulshreshth, and Christoph W. Borst. 2021. *Deep Learning on Eye Gaze Data to Classify Student Distraction Level in an Educational VR Environment.* https://doi.org/10.2312/egve.20211326

[3] Stephan Beck, André Kunert, Alexander Kulik, and Bernd Froehlich. 2013. Immersive Group-to-Group Telepresence. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (April 2013), 616–625. https://doi.org/10.1109/TVCG.2013.33

[4] Christoph W. Borst, Nicholas G. Lipari, and Jason W. Woodworth. 2018. Teacher-Guided Educational VR: Assessment of Live and Prerecorded Teachers Guiding Virtual Field Trips. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Reutlingen, 467–474. https://doi.org/10.1109/VR.2018.8448286

[5] Efe Bozkir, Philipp Stark, Hong Gao, Lisa Hasenbein, Jens-Uwe Hahn, Enkelejda Kasneci, and Richard Gollner. 2021. Exploiting Object-of-Interest Information to Understand Attention in VR Classrooms. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, Lisboa, Portugal, 597–605. https://doi.org/10.1109/VR50410.2021.00085

[6] R. Brooks. 1986. A robust layered control system for a mobile robot. *IEEE Journal on Robotics and Automation* 2, 1 (March 1986), 14–23. https://doi.org/10.1109/JRA.1986.1087032

[7] Whitney Cade, Jessica Copeland, Natalie Person, and Sidney D'Mello. 2008. Dialogue Modes in Expert Tutoring. 470–479. https://doi.org/10.1007/978-3-540-69132-7_50

[8] Yuanzhi Cao, Xun Qian, Tianyi Wang, Rachel Lee, Ke Huo, and Karthik Ramani. 2020. An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376688

[9] Ching-Yi Chang, Han-Yu Sung, Jong-Long Guo, Bieng-Yi Chang, and Fan-Ray Kuo. 2019. Effects of spherical video-based virtual reality on nursing students' learning performance in childbirth education training. *Interactive Learning Environments* (Sept. 2019), 1–17. https://doi.org/10.1080/10494820.2019.1661854

[10] Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. 2021. Towards gaze-based prediction of the intent to interact in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications (ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3448018.3458008

[11] Elena Di Lascio, Shkurta Gashi, and Silvia Santini. 2018. Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–21. https://doi.org/10.1145/3264913

[12] Sidney D'mello and Art Graesser. 2012. AutoTutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems* 2, 4 (Dec. 2012), 1–39. https://doi.org/10.1145/2395123.2395128

[13] Sidney D'Mello. 2010. Mining Collaborative Patterns in Tutorial Dialogues. 2, 1 (2010), 37.

[14] Sam Ekong, Christoph W. Borst, Jason Woodworth, and Terrence L. Chambers. 2016. Teacher-Student VR Telepresence with Networked Depth Camera Mesh and Heterogeneous Displays. In *Advances in Visual Computing (Lecture Notes in Computer Science)*, George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Fatih Porikli, Sandra Skaff, Alireza Entezari, Jianyuan Min, Daisuke Iwai, Amela Sadagic, Carlos Scheidegger, and Tobias Isenberg (Eds.). Springer International Publishing, Cham, 246–258. https://doi.org/10.1007/978-3-319-50832-0_24

[15] Cynthia D Fisherl. 1993. Boredom at work: A neglected concept. *Human Relations* 46, 3 (1993), 395–417.

[16] Kate Forbes-Riley and Diane Litman. 2011. When does disengagement correlate with learning in spoken dialog computer tutoring?. In *International Conference on Artificial Intelligence in Education*. Springer, 81–89.

[17] Hong Gao, Efe Bozkir, Lisa Hasenbein, Jens-Uwe Hahn, Richard Göllner, and Enkelejda Kasneci. 2021. Digital Transformations of Classrooms in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–10. https://doi.org/10.1145/3411764.3445596

[18] Hande Gorucu-Coskuner, Ezgi Atik, and Tulin Taner. 2020. Comparison of Live-Video and Video Demonstration Methods in Clinical Orthodontics Education. *Journal of Dental Education* 84, 1 (Jan. 2020), 44–50. https://doi.org/10.21815/JDE.019.161

[19] Yu Han, Yu Miao, Jie Lu, Mei Guo, and Yi Xiao. 2022. Exploring Intervention Strategies for Distracted Students in VR Classrooms. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–7. https://doi.org/10.1145/3491101.3519627

[20] Zhenyi He, Ruofei Du, and Ken Perlin. 2020. CollaboVR: A Reconfigurable Framework for Creative Collaboration in Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 542–554. https://doi.org/10.1109/ISMAR50242.2020.00082 ISSN: 1554-7868.

[21] Elizabeth K. Herron, Kelly Powers, Lauren Mullen, and Brandi Burkhart. 2019. Effect of case study versus video simulation on nursing students' satisfaction, self-confidence, and knowledge: A quasi-experimental study. *Nurse Education Today* 79 (Aug. 2019), 129–134. https://doi.org/10.1016/j.nedt.2019.05.015

[22] Greg Kestin, Kelly Miller, Logan S. McCarty, Kristina Callaghan, and Louis Deslauriers. 2020. Comparing the effectiveness of online versus live lecture demonstrations. *Physical Review Physics Education Research* 16, 1 (Jan. 2020), 013101. https://doi.org/10.1103/PhysRevPhysEducRes.16.013101

[23] A. Khokhar, A. Yoshimura, and C. W. Borst. 2019. Pedagogical Agent Responsive to Eye Tracking in Educational VR. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Poster*.

[24] Reed W. Larson and Maryse H. Richards. 1991. Boredom in the Middle School Years: Blaming Schools versus Blaming Students. *American Journal of Education* 99, 4 (1991), 418–443. https://doi.org/10.1086/443992 arXiv:https://doi.org/10.1086/443992

[25] Bruce M. McLaren, Krista E. DeLeeuw, and Richard E. Mayer. 2011. Polite web-based intelligent tutors: Can they improve learning in classrooms? *Computers & Education* 56, 3 (April 2011), 574–584. https://doi.org/10.1016/j.compedu.2010.09.019

[26] Bruce M. McLaren, Krista E. DeLeeuw, and Richard E. Mayer. 2011. A politeness effect in learning with web-based intelligent tutors. *International Journal of Human-Computer Studies* 69, 1 (Jan. 2011), 70–79. https://doi.org/10.1016/j.ijhcs.2010.09.001

[27] Hugh Mehan. 2013. *Learning Lessons: Social Organization in the Classroom*. Harvard University Press. https://doi.org/10.4159/harvard.9780674420106 Publication Title: Learning Lessons.

[28] Cuong Nguyen and Feng Liu. 2016. Gaze-Based Notetaking for Learning from Lecture Videos. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 2093–2097. https://doi.org/10.1145/2858036.2858137

[29] Reinhard Pekrun, Thomas Goetz, Lia M Daniels, Robert H Stupnisky, and Raymond P Perry. 2010. Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology* 102, 3 (2010), 531.

[30] Ian H Robertson, Tom Manly, Jackie Andrade, Bart T Baddeley, and Jenny Yiend. 1997. Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia* 35, 6 (1997), 747–758.

[31] Claudia Roda and Julie Thomas. 2006. Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior* 22, 4 (2006), 557–587.

[32] Mathieu Rodrigue, Jungah Son, Barry Giesbrecht, Matthew Turk, and Tobias Höllerer. 2015. Spatio-Temporal Detection of Divided Attention in Reading Applications Using EEG and Eye Tracking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, Atlanta Georgia USA, 121–125. https://doi.org/10.1145/2678025.2701382

[33] Maura Sabatos-DeVito, Sarah E Schipul, John C Bulluck, Aysenil Belger, and Grace T Baranek. 2016. Eye tracking reveals impaired attentional disengagement associated with sensory response patterns in children with autism. *Journal of autism and developmental disorders* 46, 4 (2016), 1319–1333.

[34] Ismael Santos, Peter Dam, Pedro Arantes, Alberto Raposo, and Luciano Soares. 2016. Simulation training in oil platforms. In *2016 XVIII Symposium on Virtual and Augmented Reality (SVR)*. IEEE, 47–53.

[35] Kyoungwon Seo, Samuel Dodson, Negar M. Harandi, Nathan Roberson, Sidney Fels, and Ido Roll. 2021. Active learning with online video: The impact of learning context on engagement. *Computers & Education* 165 (May 2021), 104132. https://doi.org/10.1016/j.compedu.2021.104132

[36] Jonathan Smallwood, Merrill McSpadden, and Jonathan W Schooler. 2008. When attention matters: The curious incident of the wandering mind. *Memory & Cognition* 36, 6 (2008), 1144–1150.

[37] Santawat Thanyadit, Parinya Punpongsanon, Thammathip Piumsomboon, and Ting-Chuen Pong. 2022. XR-LIVE: Enhancing Asynchronous Shared-Space Demonstrations with Spatial-temporal Assistive Toolsets for Effective Learning in Immersive Virtual Laboratories. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (April 2022), 136:1–136:23. https://doi.org/10.1145/3512983

[38] Santawat Thanyadit, Parinya Punpongsanon, and Ting-Chuen Pong. 2019. ObserVAR: Visualization System for Observing Virtual Reality Users using Augmented Reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 258–268. https://doi.org/10.1109/ISMAR.2019.00023 ISSN: 1554-7868.

[39] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Björn Hartmann. 2019. TutoriVR: A Video-Based Tutorial System for Design Applications in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300514

[40] Frank Ulrich, Niels Henrik Helms, Uffe Poulsgaard Frandsen, and Anne Vollen Rafn. 2021. Learning effectiveness of 360° video: experiences from a controlled experiment in healthcare education. *Interactive Learning Environments* 29, 1 (Jan. 2021), 98–111. https://doi.org/10.1080/10494820.2019.1579234

[41] Idalis Villanueva, Brett Campbell, Adam Raikes, Suzanne Jones, and LeAnn Putney. 2018. A Multimodal Exploration of Engineering Students' Emotions and Electrodermal Activity in Design Activities: A Multimodal Exploration of Engineering Students' Emotions. *Journal of Engineering Education* 107 (Sept. 2018). https://doi.org/10.1002/jee.20225

[42] Jason W. Woodworth and Christoph W. Borst. 2017. Design of a practical TV interface for teacher-guided VR field trips. In *2017 IEEE 3rd Workshop on Everyday Virtual Reality (WEVR)*. IEEE, Los Angeles, CA, USA, 1–6. https://doi.org/10.1109/WEVR.2017.7957713

[43] A. Yoshimura, A. Khokhar, and C. W. Borst. 2019. Eye-gaze-triggered visual cues to restore attention in educational VR. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR), poster*.

[44] Carmen Zahn, Roy Pea, Friedrich W. Hesse, and Joe Rosen. 2010. Comparing Simple and Advanced Video Tools as Supports for Complex Collaborative Design Processes. *Journal of the Learning Sciences* 19, 3 (July 2010), 403–440. https://doi.org/10.1080/10508401003708399