



Discipline vs guidance: comparison of visual engagement approaches in immersive virtual environments

Hyeopwoo Lee¹ · Jinki Jung² · Heung-Kyu Lee¹ · Hyun Seung Yang¹

Received: 15 May 2020 / Revised: 30 October 2020 / Accepted: 9 December 2020 /

Published online: 10 January 2021

© Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Immersive virtual environments (IVEs) have been extensively investigated for applications in education and man-power training because of the benefits of immersion-driven experiences as immersion becomes a factor that can both accelerate and hamper learning depending on the user's area of focus, which supports the importance of engagement. In this paper, two fundamental approaches to visual engagement in IVE are compared: discipline and guidance. The approaches aim to foster the learner's engagement to predefined area to be focused by either subtracting visual stimuli (discipline) or appending visual indicators pointing to the area (guidance). The experimental results showed no significant improvement in memory recall accuracy and time. However, the guidance group showed superior performances in usability metrics. Interestingly, a significant difference was found in the objective measure of the participants' gaze pattern revealing that the discipline makes the user's gaze consistent and stable.

Keywords Virtual reality · Immersive virtual environment · Virtual training · Learning engagement · Visual engagement · Discipline · Guidance

1 Introduction

Technological advances in the fields of virtual reality (VR) and augmented reality (AR) hardware have resulted in transformational changes while also introducing educational, social, and economic benefits to industries that have taken the lead in their adoption [38]. In particular, the fields of education and training have focused on investigating the distinctive features of AR and VR, that is, full immersion, for the enhancement of learning transfer by

✉ Hyeopwoo Lee
leehyeopwoo@kaist.ac.kr

Jinki Jung
your.jinki.jung@gmail.com

¹ School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

² Digital Maritime Consultancy Aps, H.C. Ørstedsvej 13, 1 tv, 1879 Frederiksberg, Denmark

engaging the user in immersive virtual environments (IVEs) [11, 27, 42]. Thus, the usability and fidelity of contemporary VR experiences have been exploited [25, 33]. Full immersion in IVEs helps the user develop a fine-grained mental representation of the educational (mostly spatial) subject [39] while encouraging the presence that fosters learning transfer [23, 41]. In [20], the authors experimentally reported that the existence of objects created by visual stimuli that captures a user's attention can boost learning transfer. Visual stimuli can also be used to induce a user's engagement in learning [28]. The full degree of freedom of perspective in IVEs, however, can be obstructive to the learning transfer process due to the magnificent amount of cognitive load [22, 26], which highlights the importance of engagement in learning [10]. In order to make the engagement measurable we decompose an IVE into two areas; an area dedicated to learning, i.e., informative area, and the other ambient area for building immersion. The definition of engagement in this study is then formulated as engagement with the informative area that fosters the primary objective of a program, learning. One example of such learning environments requiring a user's focus toward specific area is driving; this task demands the most perception of the driving context from a frontal view, as many scientific approaches have been taken to bring the user's attention toward the front [29, 35]. Similar to the case of driving, a user in an IVE needs to pay attention to a specific area or object, which can either be static or dynamic, to perceive the information to be transferred.

Two questions we address in this paper are the following. (1) "How can learning in IVE attract user's engagement to areas in which the information is visually presented?" (2) "How does the visual engagement system for that purpose affect user's performance and usability?" To answer these questions, our study starts by setting two fundamental approaches on visual engagement, discipline and guidance, by applying the classic learning policies in terms of correcting learners' distraction patterns. The inspiration is given from two canonical strategies, punishment and reward [6, 13] for effective education with social considerations. In an educational environment a teacher encourages students to behave properly in the social context through punishment, reward, or the combination of both. The problem of building user's engagement pattern with the informative area from given visual stimuli can take these approaches. We define discipline as a method of penalizing distraction, i.e., focusing the outside of the informative area, by eliminating the visual stimuli from user's sight, eventually a blank screen. We define guidance as a method of educating the user to correct the distraction by providing indicators pointing the informative area. These approaches can be considered as the essence of adaptive guidance [3], which assists learners in making effective learning decisions. Throughout the experiment, the effects of both approaches are investigated with respect to how they affect the mental model of learning engagement [16] and, thus, the user performance [8], usability, and gaze patterns.

2 Related works

2.1 Learning in IVEs

As visual stimuli plays a major role in human perception IVE promotes the illusion of the real world by developing a presence from a 3D scene observed through a head-mounted display(HMD) [23]. The benefit of using IVEs is magnified when it comes to learning of spatial information. Waller et al. demonstrated that the effectiveness of learning in IVEs is similar to that of real world training with long exposure [39]. Recently, the use of IVEs has expanded to the learning of motor skills [8], motor rehabilitation [14], and professional

skills [17]. Jung and Ahn described the attractiveness of virtual training to various industries because of its explicit cost-effectiveness and enhanced, safety considerations [17]. Although the higher fidelity of IVEs is expected to accelerate positive aspects of experiences in the virtual world, Mania et al. reported that the environment with lower fidelity (flat shaded) paradoxically helped the participants to be more aware of the visual identity of the recognized object based on mental images [24]. Makransky et al. reported the information overload and distraction of the learner from the use of immersive virtual reality resulting less learning outcomes in a science lab simulation [22], which supports the importance of engagement in learning at IVE.

2.2 Attention in VR

Most of the previous works in the field of VR on attention, a selection mechanism of humans that can be driven by the scene (bottom-up) or expectation (top-down) [4], have exploited the context of treatment for individuals with cognitive or behavioral impairments [7, 26, 40]. Cho et al. demonstrated the applicability of VR as a tool for enhancing attention in education [7]. A wide range of surveys on VR for pediatric neurorehabilitation, which were performed by Wang and Reid [40], suggest that interactivity and feedback play an important role in inducing active engagement and reinforcement. Multi-modality VR, such as haptic and gestures, demonstrates the capability to achieve a more realistic interaction, thus validating its suitability as a rehabilitation tool [14]. Beeharee et al. reported that the use of attention in VR can enable the efficient management of distributed virtual environments when the expected limited network bandwidth in [2] considered. Godse et al. demonstrated that the size of visually perceived objects which affects the attention of a user, enhances a user's performance when it is large and decreases when it is small [12].

2.3 Visual attention interface in VR

It is recognized that, the earlier versions of HMDs, there was a need to guide users' attention at the level of head pose owing to limitations in the field of view (FoV) [5, 30]. As the fidelity available in contemporary VR devices has increased, the level of details of attention has been evolved to gaze. Danieau et al. demonstrated four visual effects to intentionally drag a user's intention in immersive VR [9]. The authors reported a tradeoff between the efficiency and the visibility of the effects, implying that disturbing effects may attract the user better than implicit effects although the effects hamper immersion. The most supportive study conducted by Nielsen et al. [28] addressed the problem of guiding attention in cinematic VR, which provides full immersion to scenes consisting of diegetic and non-diegetic elements. The paper compares two groups of attention guidance: a forced rotation group controlling a seat over and a virtual firefly attracting the user's gaze. The firefly group reported a better experience in the presence of the forced rotation group.

3 Material and methods

This study was designed to analyze the effect of visual engagement approaches on learning in the IVE. We first describe the problem and its background in detail and then introduce two approaches used in our experiment. Finally the hypotheses for the experiment are given.

3.1 Problem definition

This paper deals with the problem of inducing visual engagement of a user to a specific area from the stimulus given from IVE. We define the term, informative area, a part of IVE directly related to information to be transferred throughout a training program. Distraction from the informative area is interpreted as that from the information transfer and thus training. In order to quantify the distraction we define the term, immersive area, which is a seductive detail (i.e., interesting but irrelevant material) of IVE. The problem in visual engagement is to determine how the system can support the learner's engagement with the informative area. The visual engagement system should be able to detect the distraction of the user and provide feedback to reclaim the user's attention, so the user adheres to the purpose of the program.

3.2 Informative and immersive area

Distinction of informative area and immersive area is an essential prerequisite to identifying the user's engagement in learning from the experimental setup, while the user's first-person perspective does not reveal the differences, as both take the form of virtual objects and effects. In the interactive training program in an IVE, the virtual objects belong to the informative area when those are connected to the flow of a training scenario. For example, in the claim of a program that teaches how to operate a fire extinguisher in a fire drill, the fire extinguisher is identified as informative area as it is a major subject to be manipulated, and other virtual objects such as fire effects are identified as the immersive area if those are used to support building immersion but not the narrative of a scenario. The distinction becomes more vivid when it comes to the perspective of its implementation. In other words there exists reusability of informative area where it enables to reuse of the same scenario with different background, i.e., immersive area. Back to the fire drill example, even though the immersive area and the scale and location of informative area can be changed, the major flow of the scenario would not be hampered by those changes.

3.3 Visual engagement approaches

To induce the user's visual engagement, it is necessary to measure the user's engagement state. Based on these measurements, visual engagement system is activated when the distraction is observed, helping the user to refocus on informative area. The system is deactivated when the user restores the attention. This flow is illustrated in Fig. 1. In the pipeline, we determine whether an user is distracted based on the sensor's measurement value. When the measurement reveals distraction, the system, which would follow either discipline or guidance approaches, will be activated.

In our experiment the eye gaze tracker and head tracker of the HMD facilitate the measurement of the engagement state in real time. It is based on a strong assumption of the productivity of training being proportional to the duration of engagement, which is not always true in some contexts [1]. When it is determined that the user's attention is needed, the visual engagement system stimulates the user visually.

As Makransky et al. noted in [22], the extraneous processing caused by the perceptual realism of high immersions from IVE can create massive cognitive overloads which contributes poor learning performance. The visual engagement system is required to lead the user's engagement from its intensive cognitive process. Discipline recovers user's focus by reducing the cognitive loads from the visual stimuli of IVE and then awakening the user's

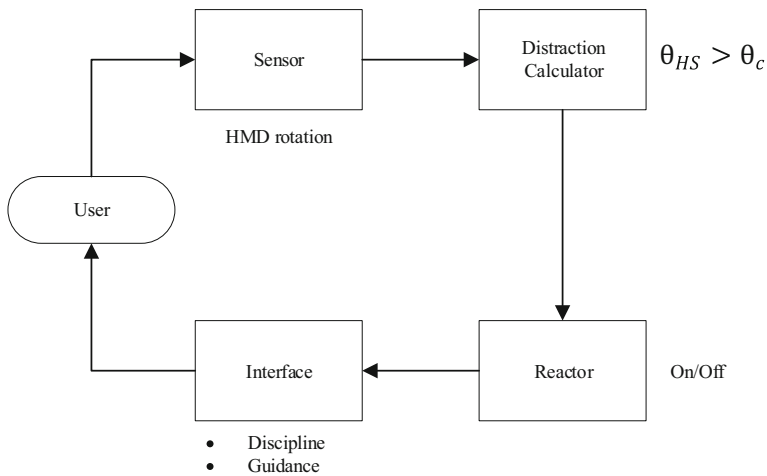


Fig. 1 Pipeline of the visual engagement system

meta-cognition. Guidance on the other hand uses additional information for refocusing on top of the cognitive process, which will contribute another cognitive load to be paid. The two approaches are not mutually exclusive, and each has a relationship with the cognitive loads from experiences in IVE. Both approaches also have a relationship with immersiveness. In case of the discipline, the absence of visual stimuli from IVE can lead to a decrease in the sense of presence [32], as it can be interpreted as a punishment while the subtle visual indicators of the guidance preventing a sudden break of cognitive process can be perceived as a positive reward contrasting to the discipline, although it does not provide any actual rewards.

3.3.1 Discipline

Discipline can be seen as a punishment, in terms of providing mental stress through an extreme visual change to void, which will impact to user experiences. The discipline can simply be likened to blinkers or horse tacks as it limits vision to the front side (or to the informative area in the context of our study). Discipline is a condition that assists with engagement by removing all IVEs to give the punishment about losing concentration. The absence of feedback compared to the guidance merely provides information to the user that the user is currently distracted, not in which direction to move.

If the concentration is necessary, a full black screen is displayed in the user's FoV to eliminate all the shown virtual environments. If the concentration is maintained, the existing screen is restored by removing the black screen. It extremely limits the stimuli when there is a loss of attention, excluding the direction in which focus should be maintained, and when concentration is regained, all the IVEs are restored. Throughout the negative experience of distraction, the user is expected to construct a mental model of engagement on the screen.

3.3.2 Guidance

Guidance is defined as additive visual elements that lead the user's gaze toward the area to be engaged in the informative area. Most of the existing interventions are forms of visual guidance and include virtual indicators such as 3D arrows, highlight effects, and blink effects

[30, 37]. In this study, we used the default arrow-shaped interface that Steam SDK provides for simplicity.

If the user's attention has left the informative area, the existing virtual environment will remain the same, and additional virtual objects will be added to the immersive area as an interface to support the user to focus. The difference compared to the discipline comes from visual feedback indicating the direction to the informative area, forming 2D arrows.

3.4 Hypotheses

For guidance with an additional visual interface, according to the authors in [12], the objects' visual perception can affect the user's performance in virtual reality training environments.

For the discipline, while previous studies demonstrated that visual guidance [3] is directly beneficial to training and education, few direct and relevant studies have provided the subjects with the form of stimuli in the form of the discipline. Therefore, we set the hypotheses based on the mental stress theory, regarding discipline's situation as a momentary loss of vision.

Guidance shows additional objects to the user, but Discipline deletes the items, making the hypothesis a more practical in-memory task from a Cognitive load [22] perspective.

In summary, the following hypotheses are formulated:

Hypothesis 1 *The presence of the engagement interface affects the user performance (H_1).*

Hypothesis 2 *Discipline will have a better user performance than guidance (H_2).*

Hypothesis 3 *The presence of the engagement interface affects the user's concentration (H_3).*

Hypothesis 4 *Discipline will have a better user concentration than guidance (H_4).*

Hypothesis 5 *The presence of the engagement interface has a positive effect on information delivery (H_5).*

Hypothesis 6 *Discipline will have a better positive effect on information delivery (H_6).*

4 Experiment

In this study, we designed a comparative experiment on visual engagement approaches focusing on the following specificities to further enhance the comparison of the discipline and guidance. The experiment design was inspired by the following [19]:

1. There is no connection between the receiving environment and information transfer.
2. The informative and immersive areas use only visual senses.
3. A screen as an informative area is placed within the virtual environment.
4. Distraction is observed as gaze patterns moving out from the virtual screen, that is the informative area.

The IVE used in the experiment possesses an immersive area that plays the role of distraction and a virtual screen as the informative area that conveys the information to learner (Fig. 2).



Fig. 2 Informative and immersive area from the learner's perspective. The yellow rectangle indicates the informative area where the video lecture is given. The rest of area in the screen is designed as the immersive area

Users can freely turn their heads to look at the desired part and can look not only at virtual screens but also the informative area that conveys information to the learner. Further, learners can freely turn their heads to look at the desired part and can look not only on virtual screens but also at obstacles placed in various locations. If the user's attention is off the virtual screen, the visual engagement system will be triggered to restore the user's attention to the screen that requires focus in the case of discipline and guidance. In the experiment, the verification and analysis of the hypotheses were carried out through user studies in three independent tests that combined immersion and informative areas. During the experiments, we measured the user performance with respect to memory recall, the usability of the interface, and the gaze pattern.

4.1 Immersive area

The immersive area in the IVE we used was a high-quality natural forest environment to build an entirely independent environment for information transfer [36]. The natural forest environment embodied rocks, trees, grass, and land and included wind or sunlight effects to allow users to immerse themselves to the extent possible.

Owing to its monotonous configuration, we designed triggers to induce distraction, that is animated dinosaurs [34]. Fifteen animated dinosaurs were placed around the user and the screen to interfere with the view on the screen in the informative area. While the information is being transferred, the animations specified in each object can operate to distract the gaze of users, and the animations contain movements in a specific place or in a particular position.

4.1.1 Task for learning

In this experiment, we designed a task that manipulated a virtual machine with multiple types of controllers and provided lecture of the task through a video shown on a virtual

screen of the IVE following a widely used format in multimedia learning [15]. Therefore, the informative area in the experiment matches the virtual screen initially located in the learner's frontal view. The size of the virtual screen is designed to be naturally exposed to the user's engagement, as the height of the screen is almost identical to the learner's FoV when the user sits on a chair (Fig. 3). The distance between the screen and the user was 20 m, and the screen size was 10 m × 10 m.

The task consists of eight sequential sub-tasks with a similar form in the study of [21]. The spatial and appearance information of the equipment is given prior to the actual lecture as tutorial in Fig. 4). The lecture video demonstrates the operation of the equipment in eight sequential orders in a first-person view, with the descriptive subtitle of each (Fig. 5).

The subtitle explains the current sequence, the equipment to be manipulated, and the operation method in the local language. Subjects do not have to operate each device in the virtual world themselves, but they are required to retain focus and memorize how each equipment is operated in each step. The evaluation for user performance applies the score of the eight sequences with the name of the equipment and the operation method written by the user immediately after a session. To invalidate a prior knowledge, each session randomizes the equipment and its operation in each step. There are differences in the degree of manipulation for each piece of equipment, as in [18]. The lever can be set from 1 to 3 by moving the handle down, middle, or up. The button can only be pressed, while the six buttons are provided by differentiating color and its text overlay description (green, red, yellow, orange, blue, and purple). The joystick can be freely manipulated by virtually grabbing its handle, while its direction is quantified and classified into eight directions (i.e., N, S, W, E, NE, NW, SE, and SW). The wheel is a rotating object that is manipulated by its handle clockwise or counterclockwise consequently from one time to three times. In the virtual room, each piece of equipment is placed on the table and easily accessible without further stepping. The lever is on the right side of the learner, the button is on the front right side, the joystick is on the front left side, and the wheel is on the left side.

4.2 Implementing discipline and guidance

If the informative and immersive areas, which contain elements that interfere with concentration, are combined, the learner's attention can be dispersed elsewhere, rather than



Fig. 3 Image of the participant

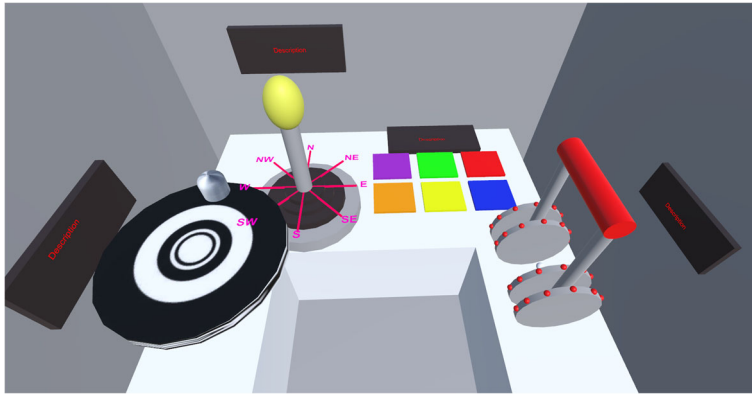


Fig. 4 Virtual machine used in the task. A wheel, a joystick, six buttons, and a lever are located on the table from left to right

focusing only on the screen. The visual engagement system should be able to make the user regain focus on the screen (Fig. 6). We checked the dispersion of the attention by calculating the angle between the user’s head and the screen. If the angle deviates from the virtual screen, the system operates such that the learner is again able to concentrate.

If raycasting is performed based on the direction vector of the learner’s HMD in the virtual environment, the concentration is determined to be dispersed, and the system is activated. If the learner turns the head 15° away from the screen, it is determined to be a dispersion of concentration.

In the discipline method, punishment is provided in the form of eliminating the visual layer from the user’s perspective. If the user’s gaze leaves the screen, the user’s vision is removed by instantly showing a black screen throughout the user’s gaze. If the learner’s eyes return to the screen with the view removed, the black screen disappears, the user’s vision is restored, and the screen is recovered. In the guidance method, when the learner’s attention leaves the screen, new virtual layers are added and shown in addition to the informative and immersive areas. In this experiment, a virtual 3D arrow interface was used [30]. To avoid interface effects, the 3D arrow indicator provided by SteamVR was used. The newly added visual layer is usually hidden, but it appears when the user’s gaze deviates from the information delivery environment screen and is displayed in four directions (i.e., up, down,



Fig. 5 Screenshots of the video used in the experiment. In the actual experiment, local language was used. **a** Tutorial video. **b** Training video

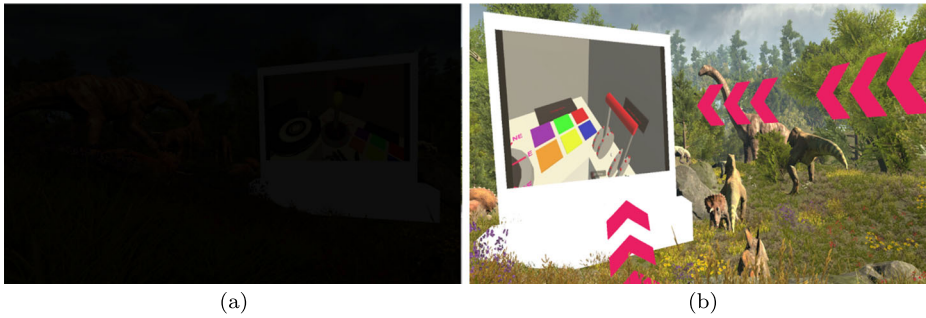


Fig. 6 Screenshots of visual engagement system. In the discipline approach, the learner's vision turns into black when requiring engagement. In the guidance method, 3D arrow shows on the learner's view to indicate the direction to focus. **a** Discipline. **b** Guidance

right, left) on the side of the screen side that requires the learner's focus. A hypothetical 3D arrow is placed in the center of view when the off-screen is based on the user. If the head is turned to the screen to enable the learner to concentrate on the screen again, the 3D arrow will be invisible. Because the display is overlaid, in the virtual environment, the informative and immersive areas remain in the user's FoV, even when the learner is not focused on the screen.

4.3 Experimental setup

There are a total of three experimental settings, all of which are combined with an informative and immersive area, the first one without any engagement system (control group), the second one with the discipline method, and the third one with the guidance method.

The subjects experienced the virtual environment by watching video clips containing information through the virtual screen. Input interfaces and sounds were not included in the virtual environment because our interest is only limited to the visual domain. A total of three tasks were experienced by the learners, and the informative and immersive areas were mixed in all the tasks. In the control group, no engagement guiding methods were employed. For the discipline method, a blackout penalty was applied when the attention left the screen. In the guidance group, the additive visual layer indicated the direction of the screen with a 3D arrow. The same person experienced these three tasks in random order, and immediately after watching the video clips containing the information twice on the virtual screen, they conducted a survey related to the usability and evaluation of the information transfer. Information from the video clips for each task is presented in Table 1. Video clips were recorded on the first-person view in the IVR with the given sequence of equipment operation with subtitles. We generated three different video clips corresponding to each task.

4.4 Measurement

Performance We measured the user's performance as a memory recall of eight sequences from the video clips containing information in terms of accuracy and time. The evaluation of the memory recall accuracy is based on how each device operates in eight sequences. After the subjects experienced the given task, they were asked to choose the correct equipment and

Table 1 Eight steps to be conducted in each group. It shows the value to be set and the type of the device at each stage. W, J, B, and L denote wheel, joystick, button, and lever, respectively

Group	Step1	Step2	Step3	Step4	Step5	Step6	Step7	Step8
Control group	B, blue	W, 2	J, south	W, 1	L, 3	B, Red	L, 2	J, southeast
Discipline group	W, 3	L, 3	B, yellow	J, southwest	L, 1	W, 2	J, west	B, orange
Guidance group	B, green	J, east	W, 1	L, 1	J, northeast	W, 2	L, 2	B, red

its operation method for each of the eight series. The total number of cases in one sequence was 20 (lever: 3, wheel: 3, joystick: 8, and button: 6). We provided the option “cannot remember” when they could not remember. The memory recall accuracy was determined by the proportion of the eight sequences that were correct. Because the primary approximation factor is the appropriate sequential information, which in turn is the primary evaluation factor, the spatial and appearance knowledge of the equipment was provided by the picture during evaluation. The memory recall time was measured during the above assessment.

Usability The usability assessments include the assessment of the interface related to engagement and a sense of immersion experienced by the learner in the virtual environment. For the sense of immersion, we asked the subjects to perform the SUS [32] questionnaire for the entire virtual environment. In the assessment of the interface, we asked them to answer a 7-point Likert scale questionnaire similar to [41], which is related to the usefulness and disturbance of the interface according to immersion and information transfer. We added questions from [41] by applying the following context of our experiment:

1. Were you able to focus on your surroundings during the task?
2. What do you think of the usefulness of information transfer during the tasks?
3. What do you think the engagement interface has done to the sense of immersion?
4. To what extent do you think the engagement interface has interfered with the transfer of information?
5. To what extent did the engagement interface help you avoid distractions?
6. Do you agree that the engagement interface was helpful to focus?

The format of the answer was a 7-point scale in increasing degree of positiveness.

Gaze pattern A gaze pattern was used to quantitatively measure the user’s concentration. The direction of the gaze was obtained from the HMD HTC Vive eye tracker. We logged the vector of the direction in every frame during the tasks. Then, we recorded the direction of the gaze in each video session separately. The following methods were used to measure the dispersion of the gaze.

1. Number of times when the learner was off the virtual screen: We counted the number of times when the gaze was off the virtual screen from all groups. For the discipline and guidance groups, this number is equal to the number of times that system was activated.
2. Measuring the dispersion of the gaze: The variance of the direction vector was calculated from the logged eye-tracking data.
3. Gaze map: We visualized the gaze vector and the 2D points projected on the frontal plane parallel to the virtual screen. The contour set and visualization of the projected 2D points are plotted.

4.5 Apparatus

The HTC Vive Eye Tracker with Tobii Eye Tracker was used to track the FoV of the subject, and the FoV of the HMD was 60. No Vive controller was used because the input interface did not exist. After experiencing the virtual environment, assessments and surveys were conducted on desktop computers. The desktop environment was Windows 10 Intel i7-8700, with 32 GB RAM and NVIDIA GeForce GTX 1080Ti, and the space for the virtual environment was set to be at least $3\text{ m} \times 3\text{ m}$. In the case of generating information videos, we recorded the video, including the first-person view in the IVE, in the same environment as in the above experiment, and the four sets of equipment were operated in eight sequences using the Vive Controller.

4.6 Participants

A total of 30 people participated in the experiment (13 females, 17 males) with an average age of 27.8 years (SD: 5.90). Here, 26 of them experienced video lectures, 26 experienced the use of AR/VR devices, and 5 experienced VR/AR device-based lecture training. In one case, the subject had a color weakness, but the task related to distinguishing the colors was performed successfully. None of the subjects experienced motion sickness during VR.

4.7 Procedure

First, the subjects watched tutorial videos on the desktop to obtain spatial knowledge and essential explanations about the information delivery environment. The tutorial video describes the location and use of the four sets of equipment. The following calibrations were then performed for eye tracking. Thereafter, the subject experienced the three tasks in random order to significantly reduce the maladjustment effect of the unfamiliar environment. A video containing sequencing information was watched on the virtual screen for 1 min in two sessions (L1 and L2), and a 30 s break was given halfway to experience the virtual environment. After the experience, the assessment and survey were conducted on the desktop. Final follow-up surveys were conducted after the last experimental environment was terminated. Owing to the COVID-19 protocols, we paid special attention to disinfection and safety whenever the experiment was performed.

5 Result

The results indicated the memory recall performance of the subjects, available through the questionnaire, and gaze pattern. We used the criteria ($p < 0.05$) for statistical significance. All statistics are presented in Tables 2 and 3.

Performance To evaluate the user performance in terms of the memory recall accuracy, we compared the number of correct answers in the eight steps of the device. One-way ANOVA analysis found no statistical significance for memory recall accuracy ($F(2, 87) = 1.59$, $p = 0.209$). In the case of memory recall time, we found no significant difference between the groups ($F(2, 87) = 1.34$, $p = 0.268$) (Fig. 7).

Usability We measured the immersiveness of users' experiences using SUS, and no significant difference was found between the groups ($F(2, 87) = 0.341$, $p = 0.712$) (Fig. 8). In

Table 2 Statistics of user performance and usability in three groups in a form of “mean (SD)”

Measurement	Observation	Group		
		Control	Discipline	Guidance
User performance	Memory recall accuracy (related to H_1, H_2)	0.850(0.209)	0.854(0.183)	0.925(0.153)
	Memory recall time (related to H_1, H_2)	54.0(43.9)	39.3(17.2)	40.2(21.8)
Usability	Level of immersion (related to H_3)	4.76(0.987)	4.55(1.09)	4.70(0.940)
	Level of focusing the environment (related to H_3)	2.77(1.19)	3.13(1.31)	4.03(1.03)
	Level of the information transfer (related to H_5)	2.90(0.845)	3.23(1.36)	3.77(1.10)
	Recognize engagement system (related to H_6)	—	5.1(1.97)	5.97(1.63)
	Obstruction of the engagement system (related to H_4)	—	4.53(2.10)	2.53(1.53)
	Obstruction of the engagement system in terms of information delivery (related to H_6)	—	4.47(2.06)	2.23(1.17)
	Obstruction of the engagement system in terms of immersion (related to H_4)	—	4.53(2.19)	3.07(1.95)
	Prevention distraction by the engagement system (related to H_4)	—	4.77(1.92)	5.1(1.63)
	Assistance of concentration by the engagement system (related to H_4)	—	3.93(1.76)	5.1(1.49)

terms of choosing the order of focusing the environment, we found that the ANOVA showed a statistically significant difference between the groups ($F(2, 87) = 9.10, p < 0.05$) After performing the Turkish HSD test, the comparison showed that the guidance group had a better concentration than the control group, the guidance group had a better concentration than the discipline group, and there was no difference between the control and the discipline groups.

Regarding the order in the information transfer overall, the result of the ANOVA showed a significant difference between the groups ($F(2, 87) = 4.56, p < 0.05$). Post hoc tests

Table 3 Statistics of the gaze pattern in three groups in a form of “mean (SD)”

Observation	Group					
	Control		Discipline		Guidance	
	L1	L2	L1	L2	L1	L2
Number of off-screen (related to H_3, H_4)	3.73(4.57)	3.17(4.19)	5.37(5.26)	2.27(3.62)	3.73(4.68)	2.83(4.19)
Variance of the gaze vector (related to H_3, H_4)	0.303(0.169)	0.345(0.194)	0.215(0.102)	0.244(0.137)	0.290(0.170)	0.334(0.216)

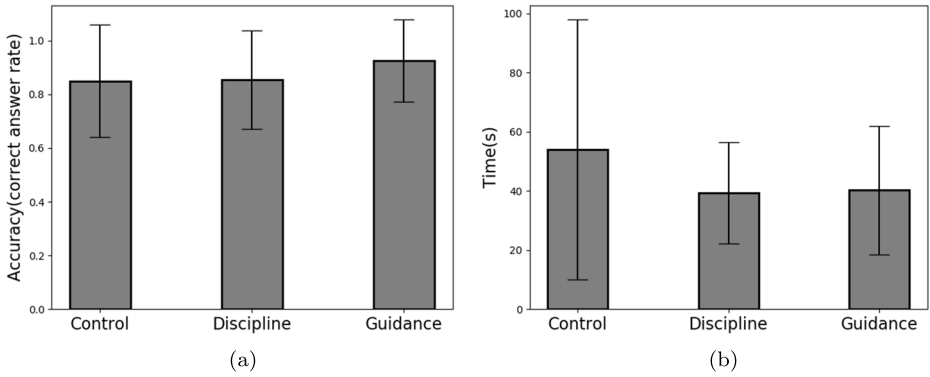


Fig. 7 Result of the performance graph **a** Memory recall accuracy **b** Memory recall time

showed that the guidance group was more useful than the control group with respect to information transfer. The graph of the usefulness score is shown in Fig. 8.

We compared the discipline and guidance groups to determine the usability.

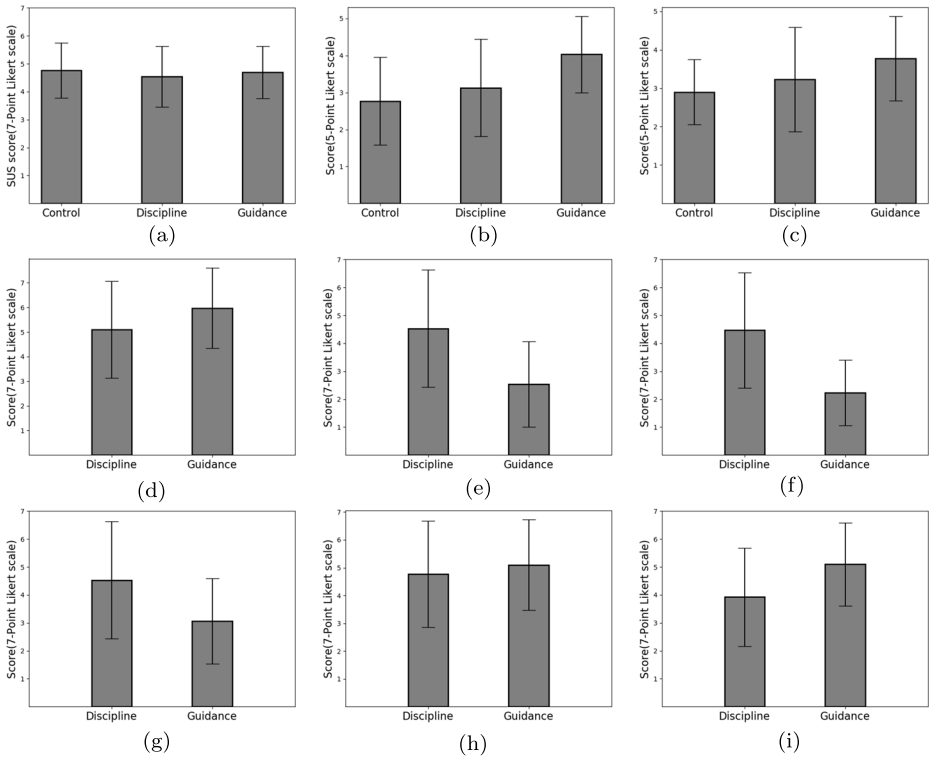


Fig. 8 Result of the usability graph **a** Level of immersion **b** Level of focusing the environment **c** Level of the information transfer **d** Recognize engagement system **e** Obstruction of the engagement system **f** Obstruction of the engagement system; information delivery **g** Obstruction of the engagement system; immersion **h** Prevention distraction by the engagement system **i** Assistance of concentration by the engagement system

In recognizing the engagement system, there was statistical significance between the two groups (discipline and guidance) ($F(1, 58) = 3.45, p < 0.05$). This implies that, in guidance, it is easier to recognize the engagement system than that in discipline. In the obstruction of the engagement system, there was a statistical significance between the two groups ($F(1, 58) = 17.85, p < 0.05$); therefore, discipline was more hindered than guidance. In the obstruction of the engagement system in terms of information, there was a statistical significance between the two groups ($F(1, 58) = 26.65, p < 0.05$). In the obstruction of the engagement system in terms of immersion, there was a statistical significant difference between the two groups ($F(1, 58) = 7.51, p < 0.05$). Therefore, discipline was more disturbed than guidance in both information delivery and immersion. In the prevention distraction by the engagement system, there was no statistical significance between the two groups ($F(1, 58) = 0.525, p = 0.236$). With the assistance of concentration by the engagement system, there was statistical significance between the two groups ($F(1, 58) = 7.66, p < 0.05$). The participants thought that guidance assists concentration rather than discipline.

Gaze pattern Two models were used to analyze the gaze pattern (Table 3, Fig. 9). In the analysis of attention, repeated measure of ANOVA with a Greenhouse-Geisser correction was used to consider the two video sessions (L1, L2). There was no statistical difference in the number of subjects losing attention to the screen between the groups ($F(2, 87) = 3.50, p = 0.868$). There was a significant difference in terms of the video sessions ($F(1, 87) = 12.8, p < 0.05$). In the case of dispersion of the gaze, statistical significance was found between the groups and sessions ($F(2, 87) = 5.58, p < 0.05$) ($F(1, 87) = 39.974, p < 0.05$). Post hoc tests showed that discipline group showed less gaze movement than the Guidance and Control groups ($F(1, 118) = 11.33, p < 0.05$, $F(1, 118) = 7.79, p < 0.05$). However, no differences were found between the Guidance and Control groups ($F(1, 118) = 0.130, p = 0.719$).

In Fig. 10, the gaze's ratio when off the screen on every frame is shown to understand losing attention to the screen deeply. We used additional visualization of the gaze vector to analyze the pattern of the gaze qualitatively. In Fig. 11, the contour set that contains the distribution of the gaze is shown to measure the gaze's rate on the view. In Fig. 13,

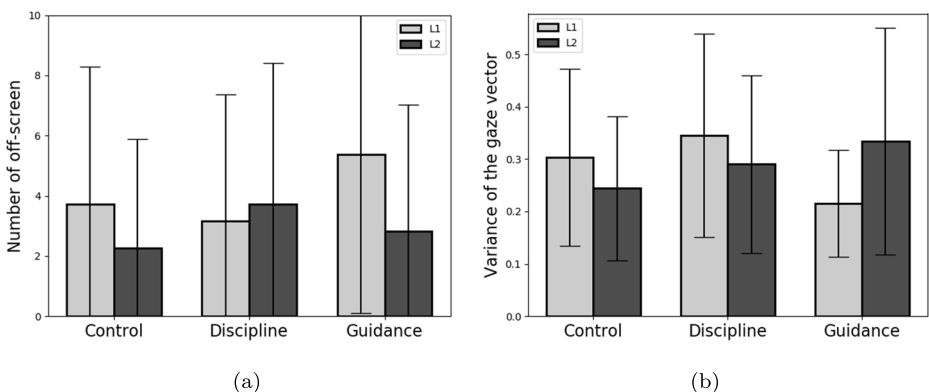


Fig. 9 Result of the gaze pattern. On **a**, number of off-screen is shown on each group. On **b**, the variance of the gaze direction vector is plotted with respect to the groups

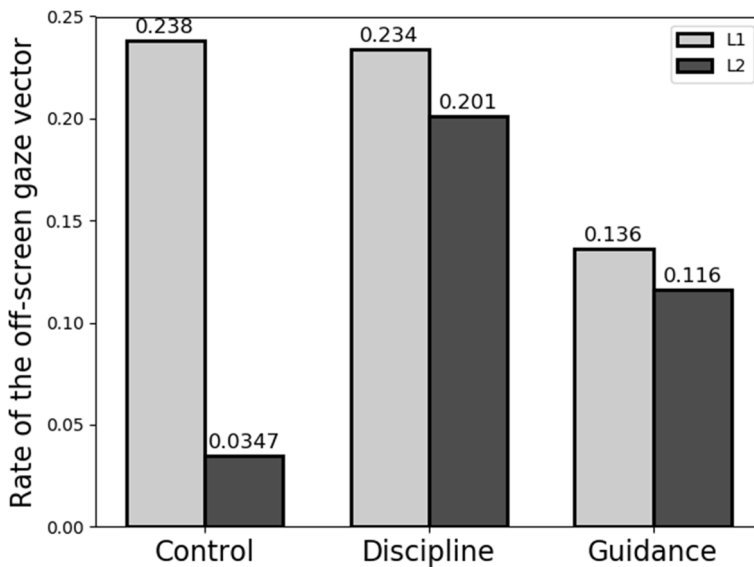


Fig. 10 Rate of the off-screen gaze vector. We counted the gaze vector when off-screen on every frame

3D histograms of the gaze pattern is revealed. All directions of the gaze around the virtual screen are depicted in Fig. 12 to see the position of the whole gaze vectors.

6 Discussion

In our experiment, the result states that both visual engagement approaches have no effect on the user performance; in other words, no statistically significant differences for both memory recall accuracy and time were found between the groups. Therefore, Hypotheses 1,2 were rejected. This result can be limited to the simplified setup and short exposure time used in the experiment. The gaze pattern showed that most of the engagement was focused on the specific position where the descriptive letters were placed in Figs. 11 and 13. We analyze this observation as the relatively small portion of the actual informative area compared to what we designed for the experiment, that is, the virtual screen. Interestingly, the control group showed almost identical performance to those with the systems, even though the clarity of the distraction pattern in the upper area of the screen, which can be interpreted as the unfocused attention [31] counteracting the cognitive exhaustive state.

In the immersion level from SUS, there was no statistical difference between the groups. The overall sense of immersion seems to be not adversely affected by the discipline and guidance approaches. Meanwhile, with respect to the effect on the user's concentration, there was a statistical difference in the level of focus on the environment. The guidance group had a better concentration than the control group. However, in the discipline group, there was no difference in the control groups (order in focusing the environment are shown in Table 2). Hence, we have confirmed Hypothesis 3 for guidance and rejected discipline in the scenario when the levels of concentration were compared.

In Hypothesis 4, the users thought that discipline was more disturbed than guidance for concentration and immersion, with statistical significance, except for prevention of

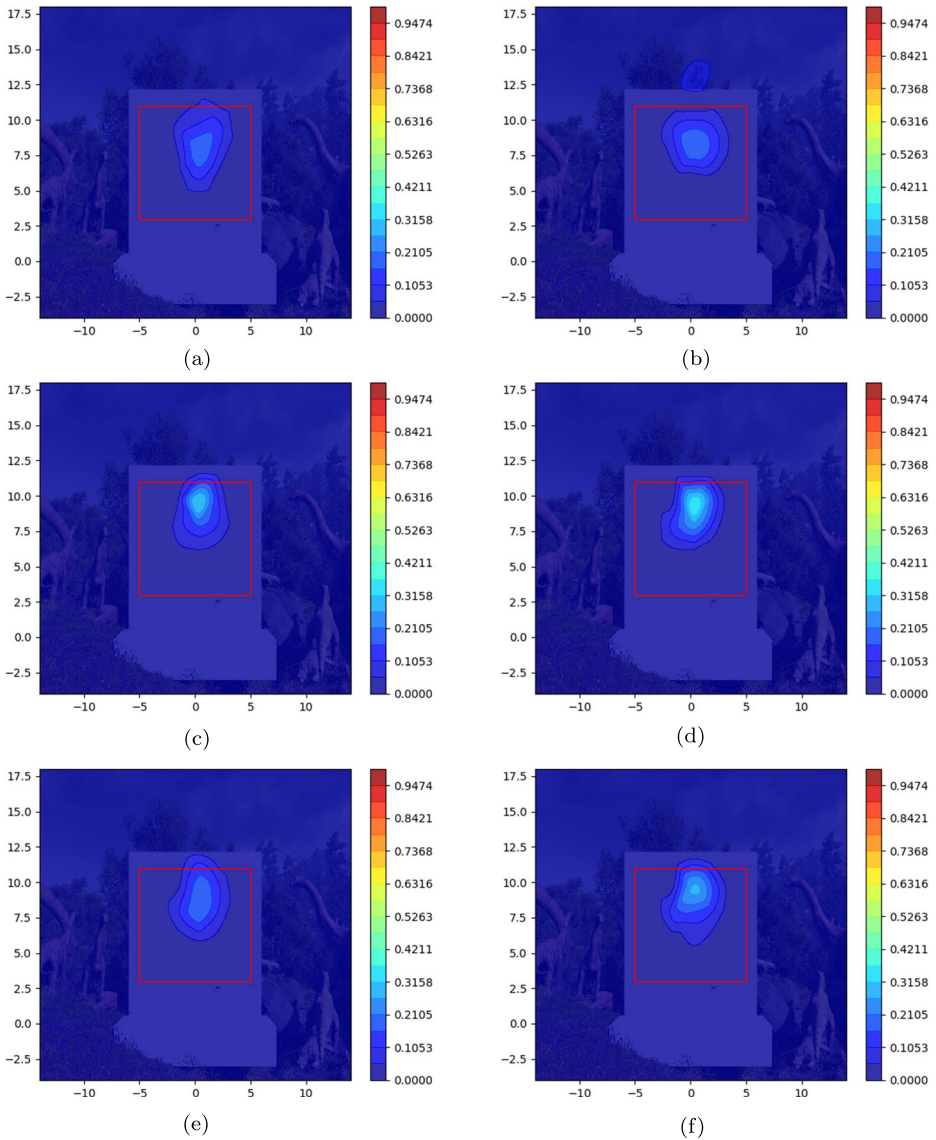


Fig. 11 Visualization of the gaze pattern with contour set. **a** Control group in L1. **b** Control group in L2. **c** Discipline group in L1. **d** Discipline group in L2. **e** Guidance group in L1. **f** Guidance group in L2

distraction by the engagement system (Table 2). However, in the result of the gaze vector in Table 3, the distribution of the gaze vector showed that the variance of the vector from the discipline was less than that of the control and guidance groups. This implies that the learners were forced to keep their eyes on the virtual screen in the discipline group, making their views fixed on the screen, and these results are converse to the survey above. Although discipline forcibly fixed their gaze compared to other environments, their subjective feelings were rather disturbing and uncomfortable.

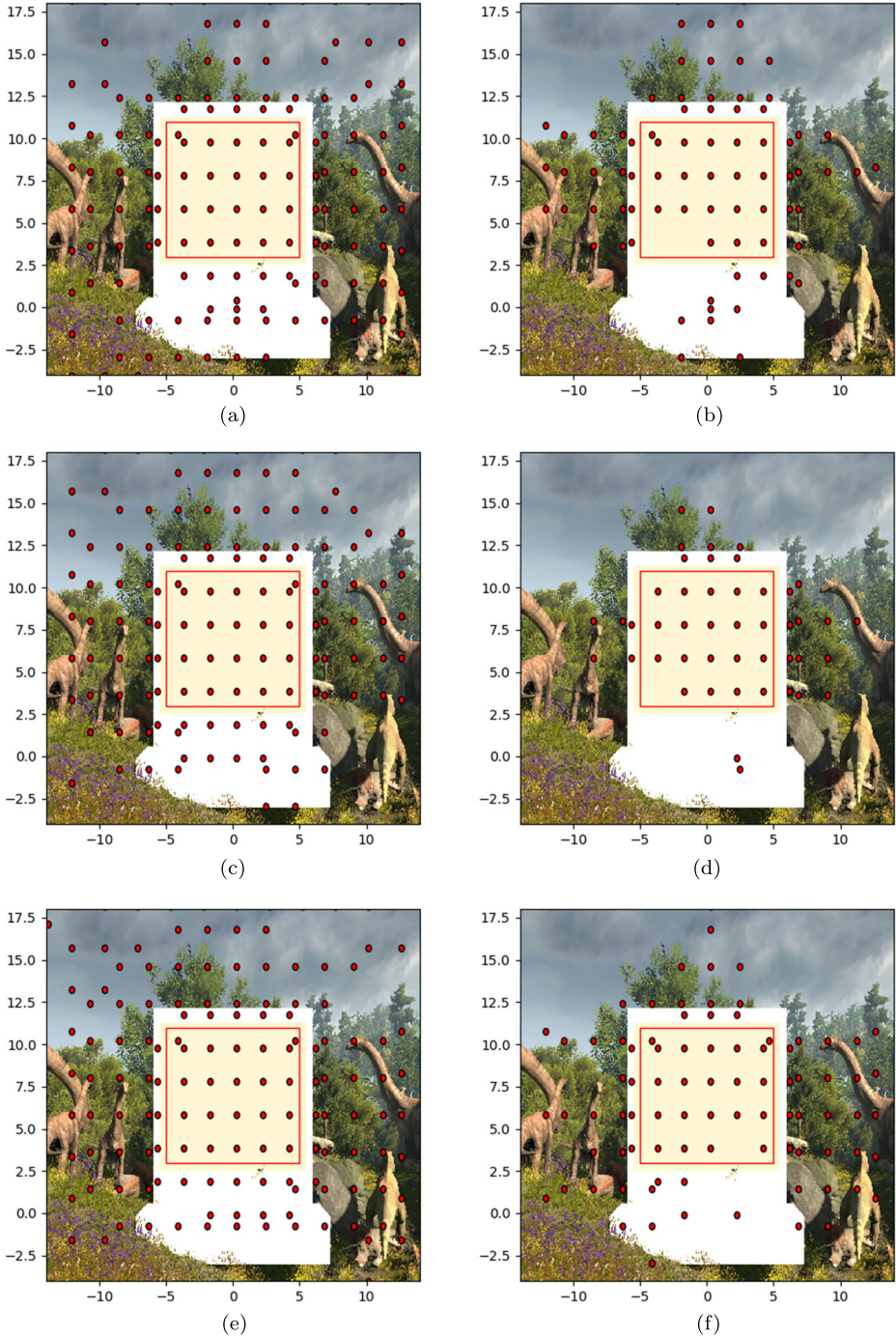


Fig. 12 Visualization of the gaze pattern with 2D projected gaze vector on the parallel plane to screen. **a** Control group in L1. **b** Control group L2. **c** Discipline group in L1. **d** Discipline group in L2. **e** Guidance group in L1. **f** Guidance group in L2

When the discipline method is applied because the learner loses focus, they are often embarrassed because the entire screen has changed without additional information. This restricts the flow of the experience, and acts as a penalty factor that undermines usability. Some users commented that they believed that it was a glitch. Unlike the discipline, the guidance approach did not make a significant difference in the display on the screen when the learner lost focus. In addition, it did not affect the user's flow of experience, and this was done by directing (using arrows) the learner's engagement in the area of information transfer. The discipline method interferes with the concentration required for information transfer. In relation to the above user experience, removing the visual layer could be interpreted as an error in the program itself, and not as an error in the user's behavior.

The results are similar to those in [28], where the use of compulsory methods to induce attention was reported to be detrimental to usability.

In Hypothesis 5, the learners thought that guidance was more useful than the Control group but not in the Discipline group, with a statistically significant difference (level of the information transfer in Table 2). Therefore, we confirmed that there is a difference between guidance and the control. Moreover, we rejected the discipline and control groups.

In Hypothesis 6, the learners observed that there was obstruction by the engagement system in terms of information delivery in discipline. For information delivery, the user performance did not differ between the groups, but the engagement interface hampered them as a subjective experience in the Discipline group. Therefore, we confirmed Hypothesis 6.

Differences in the gaze pattern among the video sessions (Table 3) indicate that visual engagement system produces slightly different effects. The actual projected gaze vector is shown in Fig. 12. The difference in the video session is that the learner is curious about the environment when watching the first video, but when watching the second video, it is determined that they had already adapted, and did not show any gaze movement. The visualization of the gaze pattern Figs. 11 and 13 shows that the gaze achieved a stationary state at the top of the screen, which was where the video's subtitles were placed.

For each group, we measured the percentage of out-of-screen in terms of the gaze vector for every frame, as shown in Fig. 10. In the case of the control group, the rates of out-of-screen visibility for L1 and L2 were similar. However, in the case of discipline and guidance, the rates of out-of-screen visibility for L2 were higher than those of L1. However, statistical analysis was not possible because it was a global calculation for all gaze vectors. Therefore, we cannot strongly argue for or against the relation between the off-screen gaze vector and the visual engagement system.

Interestingly, there were many gaze vectors just above the screen in the L2 of the Control group in Figs. 11 and 13. This is due to the absence of pre-instruction in the control group, resulting in no feedback on the exact timing. If it were a task that required more detail in concentration, the engagement attitude would have helped.

Limitations In the case of user performance, the subtitles might not have a proportional relationship between repeated attention and information delivery in the video because they were used together. Therefore, if instant information is acquired from the text in the video, engagement may not be very relevant to the performance. If there were no subtitles, it would take longer to acquire information in the video, and we believe that this would have a proportional relationship in terms of repeated focus and information transfer.

When the subjects were bored or distracted by other things, the methods for engagement were helpful (video 1 min \times 2 + a 30 s break time), and the effects were weak. Hence,

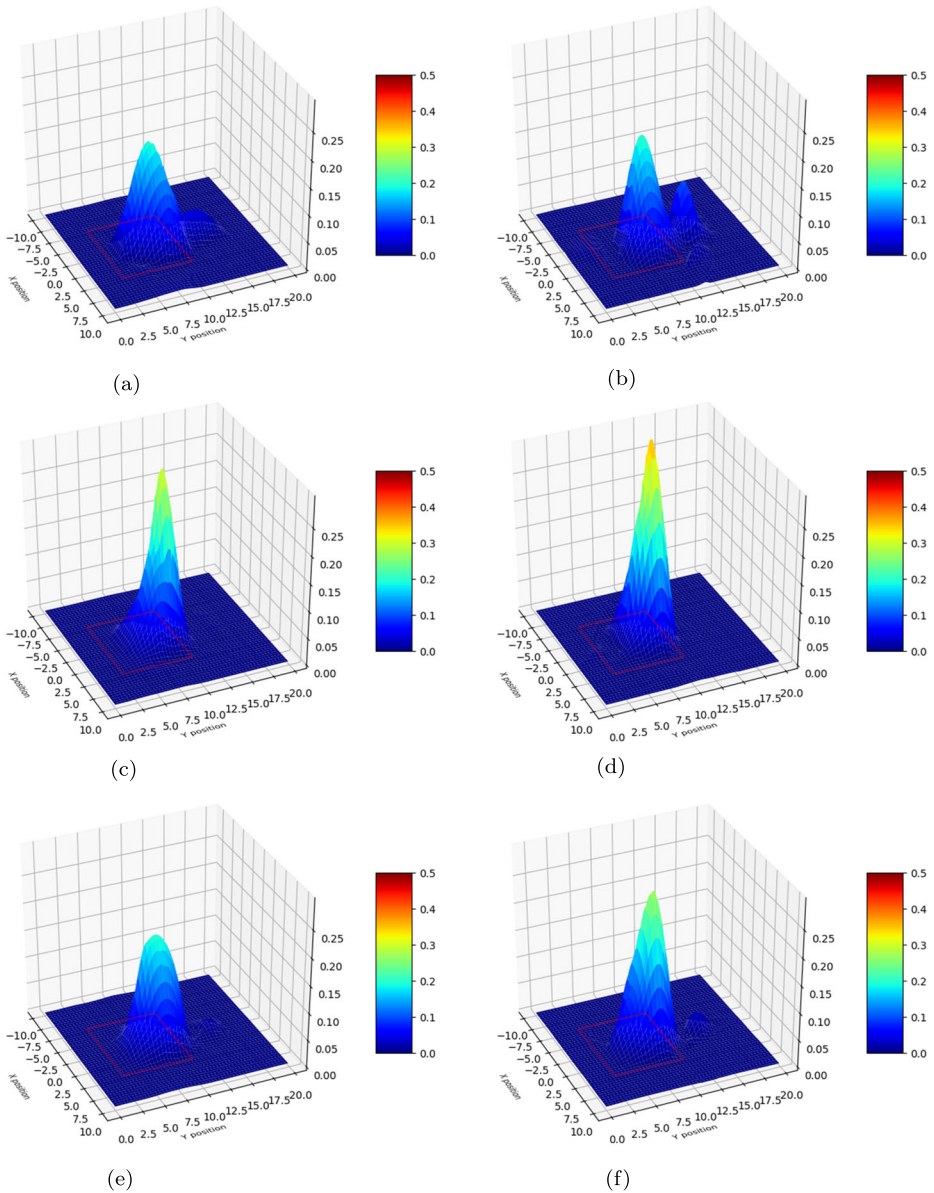


Fig. 13 Visualization of the gaze pattern with 3D histogram. Z value indicates the ratio of the gaze at the location. **a** Control group in L1. **b** Control group in L2. **c** Discipline group in L1. **d** Discipline group in L2. **e** Guidance group in L1. **f** Guidance group in L2

concentration appeared to have been less likely. The results may have been different if they were targeted at people who were less focused even after short periods.

The same person performed three independent tasks, resulting in a repetition effect despite the random order of tasks. For better experimental designs, randomly assigning subjects through A/B tests could eliminate this effect.

7 Conclusion

This study focused on the issue of engagement in IVEs. Our comparative experimental setup used discipline that eliminates visual stimuli and guidance that appends visual stimuli to guide the user's engagement. The experimental results revealed that the two ways of visual guidance do not affect user performance, as it significantly affects usability, including gaze pattern. The results confirmed that the usability of the performance of the guidance showed superior performance compared to the discipline, while the stability of the measured gaze pattern resulted in the opposite. We conclude that visual engagement approach impacts usability of learning in IVE. Designer of learning in IVE can choose one of the approaches for different purposes; the guidance approach would be beneficial when it is required to maintain the user's engagement with affordable usability while the discipline would be suitable to training of mission-critical tasks which demands intensive concentration toward informative area. It would have more impacts to the industries having the effects of two approaches for longer exposure, as it impacts to participants' behavior patterns and usability stronger. Also as future work measurement of degree of distraction and the visual engagement approaches applying that need to be investigated. The effects of the granularity of the two visual engagement approaches will then be measurable in terms of the user's performance, usability, and gaze pattern.

References

1. Baird B, Smallwood J, Mrazek MD, Kam JW, Franklin MS, Schooler JW (2012) Inspired by distraction: Mind wandering facilitates creative incubation. *Psychol Sci* 23(10):1117–1122
2. Beeharee AK, West AJ, Hubbard R (2003) Visual attention based information culling for distributed virtual environments. In: *Proceedings of the ACM symposium on virtual reality software and technology*, pp 213–222
3. Bell BS, Kozlowski SW (2002) Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Pers Psychol* 55(2):267–306
4. Borji A, Itti L (2012) State-of-the-art in visual attention modeling. *IEEE Trans Pattern Anal Mach Intell* 35(1):185–207
5. Bork F, Schnelzer C, Eck U, Navab N (2018) Towards efficient visual guidance in limited field-of-view head-mounted displays. *IEEE Trans Vis Comput Graph* 24(11):2983–2992
6. Bracey GW (1994) Reward and punishment. *The Phi Delta Kappan* 75(6):494–496
7. Cho BH, Ku J, Jang DP, Kim S, Lee YH, Kim IY, Lee JH, Kim SI (2002) The effect of virtual reality cognitive training for attention enhancement. *CyberPsychology & Behavior* 5(2):129–137
8. Covaci A, Olivier AH, Multon F (2015) Visual perspective and feedback guidance for vr free-throw training. *IEEE Comput Graph Appl* 35(5):55–65
9. Danieau F, Guillo A, Doré R (2017) Attention guidance for immersive video content in head-mounted displays. In: *2017 IEEE Virtual Reality (VR)*. IEEE, pp 205–206
10. Dede C (2009) Immersive interfaces for engagement and learning. *Science* 323(5910):66–69
11. Fernandez M (2017) Augmented-virtual reality: How to improve education systems. *Higher Learning Research Communications* 7(1):3
12. Godse A, Khadka R, Banic A (2019) Evaluation of visual perception manipulation in virtual reality training environments to improve golf performance. In: *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, pp 1807–1812
13. Heininga VE, van Roekel E, Wichers M, Oldehinkel AJ (2017) Reward and punishment learning in daily life: A replication study. *PLoS One* 12(10):e0180753
14. Holden MK (2005) Virtual environments for motor rehabilitation. *Cyberpsychology & Behavior* 8(3):187–211
15. Homer BD, Plass JL, Blake L (2008) The effects of video on cognitive load and social presence in multimedia-learning. *Comput Hum Behav* 24(3):786–797

16. Ifenthaler D, Pirnay-Dummer P, Seel NM (2007) The role of cognitive learning strategies and intellectual abilities in mental model building processes. *Technology, Instruction, Cognition & Learning* 5(4)
17. Jung J, Ahn YJ (2018) Effects of interface on procedural skill transfer in virtual training: Lifeboat launching operation study. *Computer Animation and Virtual Worlds* 29(3-4):e1812
18. Lee H, Kim H, Monteiro DV, Goh Y, Han D, Liang HN, Yang HS, Jung J (2019) Annotation vs. virtual tutor: Comparative analysis on the effectiveness of visual instructions in immersive virtual reality. In: 2019 IEEE International symposium on mixed and augmented reality (ISMAR). IEEE, pp 318–327
19. Lengenfelder J, Schultheis MT, Al-Shihabi T, Mourant R, DeLuca J (2002) Divided attention and driving: a pilot study using virtual reality technology. *The Journal of Head Trauma Rehabilitation* 17(1):26–37
20. Liang S, Choi KS, Qin J, Pang WM, Wang Q, Heng PA (2016) Improving the discrimination of hand motor imagery via virtual reality based visual guidance. *Comput Methods Programs Biomed* 132:63–74
21. Maclean P, Cahillane M (2014) Maintaining a human touch in the design of virtual parttask trainers (vppt): lessons from cognitive psychology and learning design *Proceedings of the International Conference on ICT in Education 2014*, July 3-5, Kos
22. Makransky G, Terkildsen TS, Mayer RE (2019) Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learn Instr* 60:225–236
23. Mania K, Troscianko T, Hawkes R, Chalmers A (2003) Fidelity metrics for virtual environment simulations based on spatial memory awareness states. *Presence: Teleoperators & Virtual Environments* 12(3):296–310
24. Mania K, Woolridge D, Coxon M, Robinson A (2006) The effect of visual and interaction fidelity on spatial cognition in immersive virtual environments. *IEEE Trans Vis Comput Graph* 12(3):396–404
25. Merchant Z, Goetz ET, Cifuentes L, Keeney-Kennicutt W, Davis TJ (2014) Effectiveness of virtual reality-based instruction on students' learning outcomes in k-12 and higher education: A meta-analysis. *Computers & Education* 70:29–40
26. Negut A, Jurma AM, David D (2017) Virtual-reality-based attention assessment of adhd: Clinicavr: Classroom-cpt versus a traditional continuous performance test. *Child Neuropsychol* 23(6):692–712
27. Nguyen VT, Hite R, Dang T (2019) Learners' technological acceptance of vr content development: A sequential 3-part use case study of diverse post-secondary students. *International Journal of Semantic Computing* 13(03):343–366
28. Nielsen LT, Møller MB, Hartmeyer SD, Ljung TC, Nilsson NC, Nordahl R, Serafin S (2016) Missing the point: an exploration of how to guide users' attention during cinematic virtual reality. In: *Proceedings of the 22nd ACM conference on virtual reality software and technology*, pp 229–232
29. Niezgodna M, Tarnowski A, Kruszewski M, Kamiński T (2015) Towards testing auditory–vocal interfaces and detecting distraction while driving: A comparison of eye-movement measures in the assessment of cognitive workload. *Transportation research part F: traffic psychology and behaviour* 32:23–34
30. Renner P, Pfeiffer T (2017) Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, pp 186–194
31. Rerko L, Oberauer K (2013) Focused, unfocused, and defocused information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(4):1075
32. Slater M, Usoh M, Steed A (1994) Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 3(2):130–144
33. Stevens JA, Kincaid JP et al (2015) The relationship between presence and performance in virtual simulation training. *Open Journal of Modelling and Simulation* 3(02):41
34. Studio T (2020) Dinosaur animals big pack. <https://assetstore.unity.com/packages/3d/characters/dinosaur-animals-big-pack-121763>. Accessed 2020-04-24
35. Tonnis M, Sandor C, Klinker G, Lange C, Bubb H (2005) Experimental evaluation of an augmented reality visualization for directing a car driver's attention. In: *Fourth IEEE and ACM International symposium on mixed and augmented reality (ISMAR'05)*. IEEE, pp 56–59
36. TriForge (2019) Realistic nature environment. <https://assetstore.unity.com/packages/3d/environments/realistic-nature-environment-58429>. Accessed 2020-04-24
37. Vanacken L, Grossman T, Coninx K (2007) Exploring the effects of environment density and target visibility on object selection in 3d virtual environments. In: *2007 IEEE symposium on 3D user interfaces*. IEEE
38. Vaughan N, Gabrys B, Dubey VN (2016) An overview of self-adaptive technologies within virtual reality training. *Comput Sci Res* 22:65–87
39. Waller D, Hunt E, Knapp D (1998) The transfer of spatial knowledge in virtual environment training. *Presence* 7(2):129–143

40. Wang M, Reid D (2011) Virtual reality in pediatric neurorehabilitation: attention deficit hyperactivity disorder, autism and cerebral palsy. *Neuroepidemiology* 36(1):2–18
41. Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: A presence questionnaire. *Presence* 7(3):225–240
42. Ying L, Jiong Z, Wei S, Jingchun W, Xiaopeng G (2017) Vrex: Virtual reality education expansion could help to improve the class experience (vrex platform and community for vr based education). In: 2017 IEEE Frontiers in Education Conference (FIE). IEEE, pp 1–5

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.