Can Cueing Improve Orientation in VR? The Influence of Cueing on Learning, Search Times, Mental Representation, and Presence

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Daniela Katharina Erika Decker
aus Bad Honnef

Tübingen
2025

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	16.06.2025
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Markus Huff
2. Berichterstatter/-in:	Prof. Dr. Hannes Schröter

Danksagung

Ich möchte diese Gelegenheit nutzen, mich bei einigen Personen zu bedanken, die mich in den letzten Jahren bei der Erstellung dieser Arbeit sehr unterstützt haben. Mein besonderer Dank gilt Martin, der mich von Anfang an begleitet und unterstützt hat, mir außerdem ermöglicht hat, mein Wunschthema mit Virtual Reality zu bearbeiten, und bereit war, sich mit mir all den Herausforderungen zu stellen, die dieser Bereich mit sich brachte. Ich weiß deine Unterstützung, ohne die diese Arbeit nicht möglich gewesen wäre, sehr zu schätzen. Ich durfte in diesen Jahren viel von dir lernen und hoffe, dass du genauso viel Freude an dieser Zusammenarbeit hattest wie ich.

Außerdem möchte ich mich bei Markus Huff und Hannes Schröter sowohl für die Unterstützung bei der Erstellung dieser Arbeit als auch für ihre Begutachtung dieser Arbeit bedanken. Weiterhin gilt mein Dank all meinen lieben ehemaligen Kolleginnen und Kollegen am DIE sowie meinem neuen Team an der Uni Bonn für die vielfältige Unterstützung, die mir zuteilwurde.

Besonders möchte ich mich bei Marcus Horn und Sebastian Haupt bedanken, die mich bei der Programmierung der VR-Umgebungen unermüdlich unterstützt haben. Ohne euch wären weder die VR-Umgebung noch das Design meines Pedagogical Agent möglich gewesen! Ebenso danke ich Freya Peez und Volker Schiemann, die mich maßgeblich bei der Suche nach Studienteilnehmenden unterstützt haben. Ein herzlicher Dank geht natürlich auch an all meine weiteren Freundinnen und Freunde für eure Geduld mit mir und eure Unterstützung.

Ebenfalls bedanken möchte ich mich bei meinen Eltern und Jan, die mich auf diesem Weg begleitet, unterstützt und mir so oft den Rücken freigehalten haben. Ohne euch wäre ich nicht, wo ich heute bin. Vielen Dank!

***

Table of Contents

Summary

Virtual reality (VR) offers considerable potential for learning by providing possibilities to safely and repeatedly practice educational scenarios (Arnaldi et al., 2018). However, VR can present learners with various challenges, such as how they should orient themselves in such an environment (Makransky & Petersen, 2021). Since they are surrounded by the virtual environment, they have to choose their viewing direction themselves. Therefore, learners only perceive their chosen field of view, leading to important content possibly being outside their current field of view. Moreover, in addition to this challenge, learners might also be confronted with a "dual-tasking" situation in the form of further educational content (A. D. Bender et al., 2017).

Consequently, the design of VR environments should facilitate orientation, for example, by cueing. This dissertation consists of four pre-registered experiments dealing with craft content in a virtual workshop setting. Experiment 1 (between-subject, $n = 200$) was used to assess both the credibility and appropriateness of the workshop setting used for the (desktop-) VR environments and of the so-called "pedagogical agent" used in the head-mounted display (HMD)-VR environment in Experiment 3. Learning, the perceived expertise of the agent, and the professionalism of the video were additionally investigated. In the following three experiments, I investigated the effects of different cueing methods on orientation and learning. Experiment 2.1 (within-subject design, $n = 60$) and Experiment 2.2 (between-subject design, $n = 159$), both of which were conducted online and published in one manuscript, compared light and movement cues to a condition without cues in a desktop-VR environment. The result was a consistent reduction in search times for the movement cue, which, however, entailed a reduction in presence in Experiment 2.2. The light cue reduced search times in the between-subject design in Experiment 2.2 when the type of cue was predictable, did not reduce presence and supported learning. Thus, in Experiment 3 (between-subject, $n = 91$), my intention was to test a cueing method that guided attention naturally without reducing

presence. Therefore, I compared the effects of cueing provided by a pedagogical agent's position and occasional gaze shifts to both a condition with light cueing and to a condition without cueing. Experiment 3 was a laboratory experiment with an immersive VR environment using an HMD. Both of the cueing methods in Experiment 3 reduced search times, but neither supported learning. The fact that the light cue supported learning in Experiment 2.2 but not in Experiment 3 could possibly indicate that HMD-VR is more challenging than desktop-VR. Additionally, the pedagogical agent may have been unsuccessful because it reduced presence. In summary, it's crucial to address user disorientation to fully exploit the potential of VR learning environments. My experiments showed that while various cueing methods can reduce search times, their effectiveness on learning outcomes varied. Light cues proved beneficial in predictable desktop-VR scenarios but not in HMD-VR settings, hinting at a need to develop new cueing methods.

Zusammenfassung

Virtual Reality (VR) bietet erhebliches Potenzial für das Lernen, indem es die Möglichkeit bietet, Bildungsszenarien sicher und wiederholt zu üben (Arnaldi et al., 2018). Allerdings kann eine VR-Lernumgebung für Lernende verschiedene Herausforderungen mit sich bringen, wie zum Beispiel sich zu orientieren (Makransky & Petersen, 2021). Da sie von der virtuellen Szenerie umgeben sind, müssen sie ihre Blickrichtung selbst wählen. Folglich nehmen die Lernenden nur den ausgewählten Bildausschnitt wahr, wodurch wichtige Inhalte möglicherweise außerhalb ihres aktuellen Sichtfeldes liegen könnten. Darüber hinaus werden die Lernenden neben der Orientierung möglicherweise auch mit weiteren Lerninhalten konfrontiert, was eine herausfordernde Dual-Task-Situation darstellt (A. D. Bender et al., 2017).

Daher sollte das Design von VR-Umgebungen die Orientierung erleichtern, beispielsweise durch Cueing (Bereitstellung von Hinweisen). Diese Dissertation besteht aus vier pre-registrierten Experimenten, die sich mit handwerklichen Inhalten in einem virtuellen Werkstattsetting befassen. Experiment 1 (Between-Subject-Design, $n = 200$) wurde verwendet, um die Glaubwürdigkeit und Angemessenheit dieses Werkstattsettings für die nachfolgenden (Desktop-) VR-Experimente sowie die Angemessenheit des sogenannten Pedagogical Agent (ein virtueller Instruktor), der in Experiment 3 verwendet werden sollte, sicherzustellen. Zusätzlich wurden das Lernen, die wahrgenommene Expertise des Pedagogical Agent und die Professionalität des Videos untersucht. In den nachfolgenden drei Experimenten untersuchte ich die Auswirkungen verschiedener Cueing-Methoden auf Orientierung und Lernen. Experiment 2.1 (Within-Subject-Design, $n = 60$) und Experiment 2.2 (Between-Subject-Design, $n = 159$), die beide online durchgeführt und im zweiten Manuskript veröffentlicht wurden, verglichen Licht- und Bewegungs-Cues mit einer Bedingung ohne Cues in einer Desktop-VR-Umgebung. Es zeigte sich eine konsistente Reduktion der Suchzeiten für den Bewegungs-Cue, der jedoch in Experiment 2.2 mit einer

Reduktion der Presence einherging. Der Licht-Cue reduzierte die Suchzeiten nur im Between-Subject-Design in Experiment 2.2, wo der nächstfolgende Cue vorhersehbar war, reduzierte allerdings nicht die Presence und unterstützte das Lernen.

In Experiment 3 (Between-Subject-Design, n = 91) beabsichtigte ich daher eine Cueing-Methode zu testen, die die Aufmerksamkeit automatisch lenkt, ohne die Presence zu reduzieren. Ich verglich daher die Effekte von Cueing, das durch die Position und gelegentliche Blickwechsel eines Pedagogical Agent vermittelt wurde, mit einer Bedingung mit Licht-Cues und einer Bedingung ohne Cueing in einer VR-Umgebung unter Verwendung eines Head-Mounted-Displays (HMD) im Rahmen einer Laborstudie. Beide Cueing-Methoden in Experiment 3 reduzierten die Suchzeiten, jedoch unterstützte keine den Lernerfolg. Die Tatsache, dass der Licht-Cue in Experiment 2.2 das Lernen unterstützte, jedoch nicht in Experiment 3, könnte darauf hindeuten, dass HMD-VR herausfordernder ist als Desktop-VR. Darüber hinaus könnte der Pedagogical Agent erfolglos geblieben sein, da er die wahrgenommene Präsenz reduziert hat.

Zusammenfassend ist es entscheidend, die Desorientierung der Lernenden zu adressieren, um das volle Potenzial von VR-Lernumgebungen auszuschöpfen. Meine Experimente zeigten, dass verschiedene Cueing-Methoden zwar die Suchzeiten reduzieren können, ihre Wirksamkeit auf Lernen jedoch variiert. Licht-Cues erwiesen sich in vorhersehbaren Desktop-VR-Szenarien als vorteilhaft, jedoch nicht in HMD-VR-Umgebungen, was darauf hindeuten könnte, dass die Entwicklung neuer Cueing-Methoden nötig sein könnte.

List of Publications Included in the Dissertation

a) Accepted Publications

Decker, D., & Merkt, M. (2023). Do a pedagogical agent's clothing and an animated video's setting affect learning? *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1205338

Decker, D., & Merkt, M. (2024). Shake it or light it! The effects of cueing in desktop-VR learning environments on search time and learning. *Journal of Computer Assisted Learning*, *40(3)*, 1201–1217. https://doi.org/https://doi.org/10.1111/jcal.12945

Decker, D., & Merkt, M. (2025). Look at me! Can a pedagogical agent facilitate orientation and support learning in VR? *Educational technology research and development*. https://doi.org/10.1007/s11423-025-10531-9

1 General Introduction

Virtual reality (VR), defined as a computer-generated environment (Ardiny & Khanmirza, 2018), simulating real-world or imaginary scenarios (Mikropoulos & Natsis, 2011), offers numerous educational possibilities (Arnaldi et al., 2018). One of the greatest advantages of VR in a training context is that it offers the possibility of safely and repeatedly practising learning scenarios, while reducing instructor costs (Arnaldi et al., 2018). However, VR might also challenge learners' orientation skills, especially in a 360° environment, because learners need to find the most important content themselves (Makransky et al., 2021). Thus, supporting orientation could be crucial for successful learning in VR.

Therefore, in this dissertation, I conducted four experiments. Experiment 1 was used to ensure the thematic appropriateness of the design of the setting for all upcoming (desktop-) VR experiments and the pedagogical agent for Experiment 3. The Experiments 2.1, 2.2 and 3 tested three cueing methods: cueing with light; cueing with movement; and cueing using the position and gaze shift of a pedagogical agent. The first two cueing experiments (2.1 and 2.2) compared the use of movement and light cues to a control group without cueing in a desktop-VR learning environment. Experiment 2.1 varied the independent variable cueing within subjects, and Experiment 2.2 varied the cueing between subjects. The third experiment compared cueing with the light cue, cueing provided by a pedagogical agent, and a control group without cueing in a HMD-VR environment using a between-subject design. All the cueing experiments (2.1, 2.2 & 3) dealt with the effects of these different cueing methods on the dependent variables search times and learning. Moreover, Experiments 2.1 and 3 investigated the effects on mental spatial representation as an additional focus, while Experiments 2.2 and 3 additionally examined the effects on presence and cognitive load.

The theoretical background regarding learning in VR, cueing, and pedagogical agents is presented in the next sections. Since Experiment 1 tested the design of the setting and the

thematic appropriateness of pedagogical agent's clothing, the theoretical background regarding those variables is also presented.

1.1 Learning in VR Learning Environments

Learning with VR can take place either on a PC (desktop-VR) or using a head-mounted display (HMD), which is characterised by a screen directly in front of the eyes (Buttussi & Chittaro, 2018). As the hardware required for the development and use of both systems has become more affordable (Parong & Mayer, 2018; Radianti et al., 2020), both systems have also become increasingly interesting for educational purposes. The most notable differences between the systems are that HMDs exclude visual input from the external environment, offer a larger field of view with stereoscopic vision, and precise controller tracking, whereas with desktop-VR the environment is displayed on a computer screen and thus has a smaller field of view, and the environment is controlled with the computer control elements (e.g., the keyboard) (Fromm, 2023).

The advantages of both VR systems share many similarities. Particularly in areas of training that are dangerous or costly in realistic conditions, VR offers the possibility of repeatedly practising training scenarios safely and economically. Thus, recreating those scenarios in VR may reduce risks and costs (Arnaldi et al., 2018; Renganayagalu et al., 2021). Furthermore, VR training scenarios often require learners to participate actively (Conrad et al., 2024), directly receiving feedback (Makransky et al., 2020; Mikropoulos & Natsis, 2011), which then possibly encourages the selection and integration of information (Araiza-Alba et al., 2021). Moreover, recent research implies that learners are more intrinsically motivated (Cheng & Tsai, 2019; Makransky & Lilleholt, 2018; Olmos-Raya et al., 2018) and enjoy learning with VR (O. A. Meyer et al., 2019).

However, when designing VR learning materials, certain aspects need to be considered. Because learners are surrounded by a virtual environment, finding relevant content might be more challenging (Makransky et al., 2021), and the guidance of attention could be a more

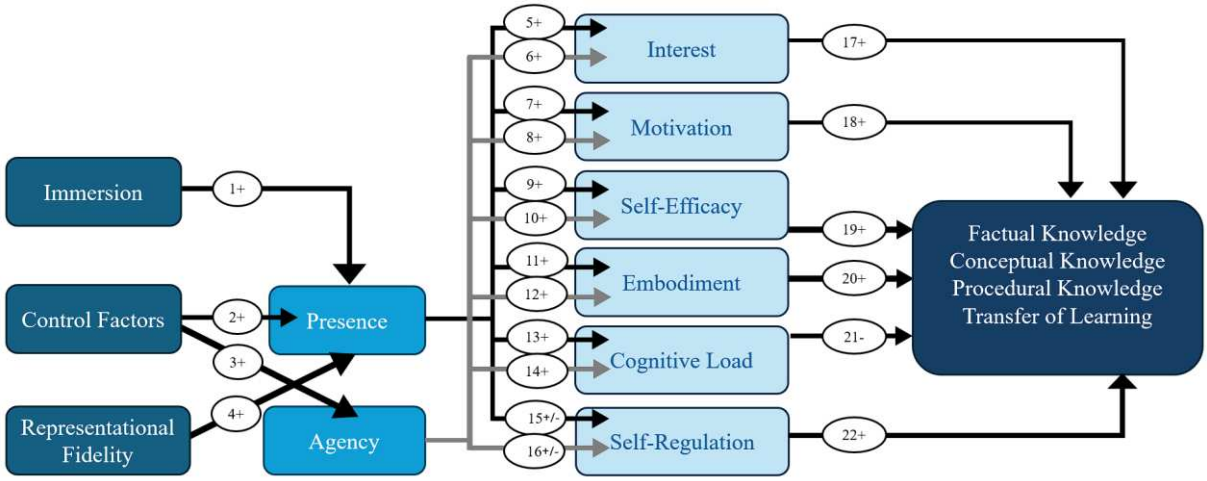
crucial factor than when using conventional learning material. Several media learning theories have explored the challenges and considerations involved in designing learning materials, such as instructional videos and virtual reality environments. These theories offer valuable insights and guidance on how to optimize the design of VR educational content. The main theories are briefly presented in the next section.

1.2 Cognitive and Social Learning Theories

According to cognitive load theory (CLT; Sweller et al., 1998), cognitive resources in working memory are limited, in contrast to the almost unlimited long-term memory capacity. According to CLT, there are three types of cognitive load: intrinsic, extraneous, and germane (Sweller et al., 1998). While intrinsic load depends on the task's complexity and the learners' prior knowledge, germane load corresponds to the effort the learners invest in constructing schemas. Extraneous cognitive load, however, occurs whenever learners must invest cognitive resources in processing information that is irrelevant for the task, possibly due to a subpar design of the learning materials or instructions (Klepsch et al., 2017; Sweller et al., 1998). Partly based on CLT (Sweller et al., 1998), and moreover recognizing that VR might pose special challenges for learners, the cognitive affective model of immersive learning (CAMIL, Makransky & Petersen, 2021) was developed. According to CAMIL (Makransky & Petersen, 2021), the instructional design in VR must uphold certain requirements for successful learning in VR so that learners can overcome the challenges that VR presents (see Figure 1).

Figure 1

Cognitive Affective Model of Immersive Learning (CAMIL, Makransky & Petersen, 2021)



CAMIL identifies three technical factors, namely immersion, control, and representational fidelity, that influence the psychological learning prerequisites – presence and agency (Makransky & Petersen, 2021). First, immersion is both a measure of the liveliness of a system and the degree it is able to exclude the external environmental input (Cummings & Bailenson, 2016). Second, the decisive factors that characterise control factors are the degree and nature of control the system allows (e.g., the extent to which the environment permits movement and the degree to which the learners' actions shape the progression of events within it) (Witmer & Singer, 1998). Third, the accuracy of representation is defined by the extent to which the representation represents the real world (Dalgarno & Lee, 2010; Makransky & Petersen, 2021). Although of these three technical factors only the control factors influence agency, which is the feeling of controlling actions (Johnson-Glenberg, 2019; Moore & Fletcher, 2012), all these technical factors influence presence. Presence is defined as the feeling of being there in a VR world (K. M. Lee, 2004; Witmer & Singer, 1998). According to K. M. Lee (2004), presence can be experienced in three dimensions, namely physical, social, and self-presence. Physical presence relates to interactions with objects and environments;

social presence involves engagement with other social actors; and self-presence reflects the individuals' perceptions of themselves in an environment.

However, even though immersion influences numerous affective and cognitive factors, such as interest and motivation, mediated via its impact on presence, recent research also implies that high levels of immersion might also lead to high levels of extraneous cognitive load (Makransky et al., 2019; O. A. Meyer et al., 2019; Parong & Mayer, 2021). One aspect that might account for an increase in extraneous cognitive load is the fact that in most VR environments, learners must choose their viewing direction themselves due to being surrounded by the virtual environment. Thus, learners only perceive their chosen and limited field of view (i.e., the maximum momentarily visible angle, Zhu et al., 2020), which varies not only depending on the learners' gaze direction but also on their position in the environment (Shi et al., 2017), possibly leading to important areas being excluded from the chosen field of view. Therefore, searching for content might increase extraneous cognitive load (Makransky & Petersen, 2021), which could be one of the reasons why VR-based learning has not necessarily led to superior learning in previous research (Leder et al., 2019; Makransky et al., 2019; Queiroz et al., 2022). An additional reason why VR-based learning could be ineffective is the increased number of potential interactive elements that might distract learners (Frederiksen et al., 2020).

Moreover, in a realistic VR learning scenario, learners may be required to orient themselves (i.e., find the most relevant field of view) while simultaneously concentrating on learning content. Thus, the challenges associated with dual tasks in VR should also be investigated, since research on dual tasking has shown that it is challenging for learners (A. D. Bender et al., 2017; Hornof et al., 2010; D. E. Meyer & Kieras, 1997; Pashler, 1994). Thus, further theoretical background regarding dual-tasks will be presented in Section 1.4.2.

1.3 Comparing Desktop-VR and HMD-VR

Regarding key aspects of the theories outlined above, it is important to highlight that there are several important differences in how desktop-VR and HMD-VR relate to certain aspects of these theories. First, extraneous cognitive load in desktop-VR may be lower than in HMD-VR since it provides less sensory input (Makransky & Petersen, 2021). In contrast, HMD-VR can increase cognitive load not only because of its higher level of sensory input but also because it requires physical movement to achieve spatial orientation. Both of these factors mean that users need to process more information, which can be both engaging and demanding (Makransky et al., 2019).

Second, regarding immersion, a central construct in CAMIL (Makransky & Petersen, 2021), it can be assumed that desktop-VR typically offers a lower level of immersion than HMD-VR. Users view the virtual environment on a flat screen, which is constrained by the size and resolution of the monitor and unable to exclude external sensory input. Consequently, the feeling of presence, defined as the sensation of being in the virtual environment, generally might be weaker in desktop-VR settings (Slater & Wilbur, 1997). By contrast, HMD-VR provides a higher level of immersion by encompassing the user's entire field of view and tracking head movements, thereby creating a more convincing illusion of being in the virtual environment. This increased immersion enhances the sense of presence and makes the experience more realistic (Shu et al., 2019; Slater & Sanchez-Vives, 2016).

Third, regarding control, which is also a key factor in CAMIL (Makransky & Petersen, 2021), there are also several differences between the two systems. Concerning the field of view, in desktop-VR, learners are required to use a keyboard and mouse to adjust their perspective, navigate the environment, and steer their actions in the environment. By contrast, the wider field of view in HMD-VR, which closely matches natural human vision, can be adjusted by head and body movements. As such, the HMD functions as an in- and output device. Further input devices for HMD-VR are motion controllers or hand tracking with

accelerometers and gyroscopes used to control the environment (Renganayagalu et al., 2021). Nevertheless, the perception of which input device is more natural and intuitive, and therefore superior in terms of control, largely depends on the learners' familiarity with the device (Seibert & Shafer, 2018).

In spite of these differences, a lack of attention guidance in both systems might cause disorientation and navigational difficulties (Marsh & Smith, 2001; Smith & Marsh, 2004), which could be an important factor influencing their effectiveness (Grogorick et al., 2018; Makransky et al., 2021). Therefore, when learners are required to simultaneously focus on learning and orienting themselves, providing orientation support could enhance the learning process. Thus, in the following sections, an overview of the current state of research on attention guidance (e.g., cueing) in VR is outlined.

1.4 Attention Guidance in VR

In visual learning materials, it is essential for learners to focus on what is relevant to ensure they receive the necessary attention for effective cognitive processing (de Koning et al., 2010). Thus, in the following sections, cueing is presented as a method to potentially assist the learners' orientation in learning material, starting with a general overview. This will be followed by a discussion of specific cueing methods that have been tested within VR environments, showcasing how these immersive settings can benefit from carefully designed visual cues.

Attention will then be given to cueing methods that have been examined in dual-task scenarios, where learners are required to manage two tasks simultaneously. Subsequently, the potential of pedagogical agents to provide cueing and effectively guide learners' attention will be examined. This begins with a general overview of research regarding pedagogical agents, followed by a deeper exploration of possible functions, such as attention guidance. Finally, the section will address the potential drawbacks of cueing, such as the possibility that learners

become overly reliant on cues, which could ultimately hinder the development of reliable mental spatial representation.

1.4.1 Cueing in VR

In general, the signalling principle emphasizes that the inclusion of cues in learning material guides the learners' attention to the most relevant parts of the learning content, assisting them in the process of information extraction and therefore facilitating learning (Alpizar et al., 2020; Boucheix et al., 2013; Richter et al., 2016; Schneider et al., 2018; van Gog, 2021). This is particularly beneficial for the first interaction with learning material (van Gog, 2021). For learning materials with a predefined field of view, such as PowerPoint presentations or learning videos, which are confined to the computer screen, the available cueing options are manifold. For example, PowerPoint animations, by using a fade-in effect to gradually reveal key points, can draw learners' focus to new information incrementally, as can the zoom function in learning videos (for further examples, see Schneider et al., 2018). These versatile cueing techniques are widely applied in both static and animated learning materials (Boucheix et al., 2013; de Koning et al., 2011; L. Lin & Atkinson, 2011). However, VR environments are characterized by the fact that learners are surrounded by a 360° environment and thus have to choose their field of view themselves (Rothe & Hußmann, 2018). Since locating highly relevant content can therefore be more challenging (Makransky et al., 2021), guiding the learners' attention could also be more difficult than in conventional materials.

Several conventional information guidance methods (e.g., the use of circles, arrows, and textual cues) have been adapted and tested for VR (Albus et al., 2021; Albus & Seufert, 2022; Zhang et al., 2023). However, given the affordances of VR, new methods have also been developed. These include forcefully taking the user to the target field of view (FOV), providing a (virtual) person or object to follow, or introducing object manipulation (i.e., changing the perceptual properties of an object). These methods vary considerably in their effectiveness and degree of obtrusiveness (Speicher et al., 2019). Specifically, regarding the

degree of obtrusiveness, it is important to distinguish between “diegetic” and “non-diegetic” cues. Non-diegetic cues are elements that are added to the scene that do not belong to the virtual world – they are overlaid on the screen to direct the users' attention. Examples of non-diegetic cues include arrows and blurring effects that are often obtrusive as they introduce external elements to the scene. By contrast, diegetic cues, for example, the intermittent switching on and off of a virtual traffic light signal or a flashing billboard, are inherently part of the scene and blend more seamlessly into the environment (Rothe & Hußmann, 2018). Contributing to this definition, Lange et al. (2020) point out that elements that are added to a scene for guidance purposes but are perceived as a part of the original scene can also be defined as diegetic cues, for example, a swarm of bees in a forest.

Regarding the effects of non-diegetic cues on learning in VR environments, positive effects on learning have been observed for textual cues (Albus et al., 2021). For instance, Albus et al. (2021) discovered that text annotations consisting of brief phrases identifying relevant parts (in this case, components of a virtual seawater desalination plant) in a text field connected to the scene element enhanced recall but did not positively impact comprehension or transfer. However, it must be emphasised that in this example the textual cues partly reflected the narration, so that “dual coding” might have occurred. As such, learners in the cued condition both heard and read the information. Therefore, the learning effectiveness might not have been solely a result of cueing. Furthermore, it was not the aim of this study to test orientation in a 360° scene, as the virtual environment was not technically set to leave the participants' field of view (Albus et al., 2021). By contrast to these partly positive results with text cues, Zhang et al. (2023) found no significant effects of textual cueing on learning effectiveness in their desktop-VR environment dealing with computer assembly. However, this may have been because in this study the textual cues partly mirrored written texts available to all learners when they clicked on an item. A further study, however, in a desktop-VR environment, used cueing implemented with green circles and found increased learning

outcomes and decreased extraneous cognitive load (Albus & Seufert, 2022). In this study, learners were asked to select several different virtual animals in a virtual forest and follow a narration about those animals. In the cued condition, green circles highlighted the part of the animal that the narration was referring to (e.g., ears and hoofs); thus, the implemented cueing guided the attention within a 3D-model learners were already attending to.

Regarding the influence of non-diegetic cues on orientation, certain non-diegetic methods, such as autopilot – which seamlessly directs users to the target area – and guiding arrows, have shown promising results in VR videos (Y.-C. Lin et al., 2017). Autopilot, however, might also cause motion sickness due to a mismatch between visual and proprioceptive input (Chang et al., 2020).

Recent research investigating the effects of non-diegetic cues on orientation combined with learning has found that using textual and arrow cues (i.e., an arrow with a text written inside stating that the learners are missing content) to direct learners' attention to parts of a VR video outside their current field of view, had positive effects on learning (Beege et al., 2023; Liu et al., 2022). One challenge in guiding attention to content beyond the current field of view lies in determining the optimal placement of cues, whether within or outside the current field of view. For example, Liu et al. (2022) compared the effects of cues that were initially in the leaners' FOV to cues that were not in the current FOV and found that cues that were inside the FOV guided attention better and resulted in better learning outcomes. However, since no interaction was necessary in these videos, several participants in conditions without cues or with cues that were initially outside their field of view did not turn around and thus watched the 360° video as they would watch a conventional video, which might have led them to miss important content. In line with this, text and arrow cues that Beege et al. (2023) implemented in the FOV also successfully guided attention to the relevant FOV and supported learning. These results underscore the need to provide guidance that helps learners focus on the most relevant areas in VR environments. While the findings suggest that cues within the

field of view tend to be more effective (Beege et al., 2023; Liu et al., 2022), alternative approaches, such as providing instructions, might also encourage learners to actively seek cues beyond their immediate field of view (Porte et al., 2024). Summing up the research regarding non-diegetic cues, none of these studies tested cueing as a way of guiding attention in an *interactive* 360° environment with relevant content possibly being outside the field of view (Albus et al., 2021; Albus & Seufert, 2022; Zhang et al., 2023).

Moreover, non-diegetic cues could reduce representational fidelity, which might negatively affect presence according to CAMIL (Makransky and Petersen, 2021). Therefore, diegetic cues, which are already a part of the VR scene, might interfere less with immersion (Lange et al., 2020) and uphold a higher level of representational fidelity, thus reaching a higher level of presence. Diegetic cues, using light and movement, which have already been analysed with spatial eye-tracking data (i.e., heatmaps), have shown promising results in directing attention in a VR movie employing a within-subject design (Rothe & Hußmann, 2018). Further, light cues enriched with spatial audio have also been used to guide the attention to speakers intending to join a conversation in a VR group, resulting in significantly faster response times than when no cues were used (G. Lee et al., 2024). Furthermore, Lange et al. (2020) developed a diegetic cueing method that seamlessly blended into the VR environment by using a virtual swarm of bees to cue which apples users had to select. They then compared this cueing method with four non-diegetic cueing methods (e.g., arrows, blurring, and circles) and showed that the diegetic method interfered the least with immersion and, consequently, presence and provided the most realistic scene. However, since diegetic cues need to be perceived as a realistic part of a scene, this specific method might not be generic enough to be used in a wide range of environments, as a swarm of bees might not be perceived as a natural part of an urban environment. A similar diegetic cueing technique was utilized by Porte et al. (2024) who demonstrated that increasing an objects' luminosity (i.e., a virtual trees' luminosity) and using marker stones in a virtual forest significantly improved the

efficiency of information gathering. In this particular virtual environment, participants were instructed to explore and select objects, such as trees and shrubs, to gather information. Upon selection, additional information about the chosen object was displayed on the screen. After completing the learning environment, participants were asked to choose a location in the forest for building a tourist attraction, prioritizing areas with minimal environmental impact and providing justifications for their choice. This task was positively influenced by the use of marker stone cues, probably because the cues during the learning phase helped participants form a more efficient and organized mental model of the forest (Porte et al., 2024). Summing up, diegetic cues offer great potential to provide guidance in realistic VR environments. However, these methods have not yet been tested with simultaneous learning tasks. As outlined above, an orientation task combined with a learning task constitutes a dual-task. As such, I will outline the theoretical background regarding cueing in dual-tasks in the next section.

1.4.2 Cueing in Dual-Tasks

Ordinarily, people don't seem to notice much difficulty performing two tasks simultaneously unless the tasks are physically incompatible or mentally demanding. However, laboratory studies combining a primary and a secondary task reveal the opposite: even simple, compatible tasks can interfere with each other significantly (Pashler, 1994). Thus, performing two tasks simultaneously seems to be challenging (A. D. Bender et al., 2017; Hornof et al., 2010; D. E. Meyer & Kieras, 1997; Pashler, 1994).

The underlying reasons for this may lie in the structure of human cognitive architecture, particularly the inherent constraints of working memory capacity. When individuals must not only retain information but also actively process it, they are typically limited to managing two to three discrete pieces of information simultaneously (Sweller et al., 1998). These limitations provide a foundation for understanding how and why dual-tasking often results in interferences. Two of the most widely accepted theories addressing these interferences are

capacity sharing and bottleneck models, both grounded in these fundamental cognitive constraints (Pashler, 1994).

Capacity sharing theories propose that the limited cognitive resources are distributed across tasks during simultaneous performance. Because these resources are finite, dividing them among tasks reduces the mental capacity available for each, leading to diminished performance in one or both tasks (Pashler, 1994; Wu & Wang, 2024). Similarly, bottleneck models align with these constraints, suggesting that certain mental operations require exclusive access to a specific mechanism, causing delays or impairments when tasks compete for it (Pashler, 1994). In line with that, impairments can also arise at the motor stage, where bottlenecks in response initiation or execution can lead to delays in dual-task performance, even when the tasks involve different effectors, such as responding vocally in one task while using manual key presses in another (Bratzke et al., 2008). In summary, these models account for observed interference by highlighting resource limitations and processing constraints (Pashler, 1994).

However, dual-task research suggests that overall performance (i.e. performance in both tasks) might be enhanced when performance in one or both tasks is facilitated, for example by training (A. D. Bender et al., 2017; Dux et al., 2009; Hazeltine et al., 2002; Strobach et al., 2012), cognitive offloading (Grinschgl et al., 2023), or instructional support (Hopkins et al., 2017; Peng et al., 2022). Regarding training, Bender et al. (2017) showed that dual-tasking performance improved for participants who trained on the combined visuomotor tracking and shape discrimination tasks (72 trials over six sessions) but not in those who trained the tasks separately (36 trials each). Importantly, this benefit was not attributable to baseline differences, as both groups exhibited comparable dual-task interference initially, and the observed training-related improvements appear to have limited transferability to other dual-tasks (A. D. Bender et al., 2017). Moreover, Hazeltine et al. (2002) highlighted the necessity

of designing instructions that encourage participants to prioritize both tasks equally when training them concurrently.

Further, Grinschgl et al. (2023) studied cognitive offloading, which is defined as the externalization of cognitive processes (e.g., taking notes) in a dual-task setting involving a primary pattern copy task (PCT), where participants recreated color patterns by dragging and dropping colored squares into a workspace, switching between a model window (to view the pattern) and a workspace window. Offloading was measured by how often participants reopened the model window to reduce memory demands instead of relying on internal working memory. Temporal costs were introduced by imposing a 2-second lockout for half of the participants each time they accessed the model window, requiring them to wait before viewing the pattern again. The secondary task was an auditory N-back task, where participants identified matching consonants two steps apart. The results showed that participants who experienced no temporal costs for offloading in the PCT performed better on the N-back task, indicating that offloading released cognitive resources. This suggests a dynamic interplay where increased offloading in the primary task improves performance in concurrent, unrelated tasks by freeing working memory capacity (Grinschgl et al., 2023).

Finally, in the context of instructional support, cueing could serve as an effective strategy for facilitating performance in a dual-task setting (Hopkins et al., 2017; Peng et al., 2022). In particular, cross-modal cueing, defined as cueing that involves different sensory modalities (e.g., using a visual arrow to cue the direction of an auditory target), has been shown to successfully support spatial tasks (Blurton et al., 2015). However, research regarding whether it might also enhance the performance of an independent secondary task remains scarce. For instance, Hopkins et al. (2017) conducted a study in which participants were asked to perform a dual-task that involved a primary visual search task and a secondary auditory task. The primary task required participants to locate a horizontal line among distractors (i.e., slightly slanted short lines) in a visual display by pressing a corresponding

key. Simultaneously, they performed a secondary task that involved listening to a series of numbers to identify any repeated pairs. For the primary visual task, Hopkins et al. (2017) then compared two types of cueing, namely auditory and tactile. The auditory cueing was implemented by delivering a 500 ms sine wave tone through speakers located 60 cm to the left and right of the main search monitor. The tactile cueing was delivered via a 500 ms vibration through tactors affixed to a belt around the participant's torso, indicating in which half of the display the target was located (i.e., the side the cue appears on indicates the side on which the target is located). Results indicated that auditory cues were not effective in a dual-task scenario, supporting neither the visual nor the auditory task, as they interfered with the secondary auditory task. This interference likely occurred because both the auditory cue and the auditory task were competing for the same attentional resources within the auditory modality. By contrast, the tactile cues successfully supported the visual task and did not seem to interfere with the auditory task. However, even though these results might suggest that tactile cues could be effective in dual-tasking, the support used in the visual task did not simultaneously enhance accuracy in the auditory dual-task. Thus, cueing support in one task using another modality (e.g., using tactile cues in a visual task) might not translate to a better performance in a secondary task that is presented in another modality (e.g., an auditory task, resulting in a total of three modalities in the dual-task).

Therefore, selecting the appropriate cueing modalities in a dual-task may introduce additional complexities, further compounding the challenges inherent in VR, which have already been outlined. However, based on the results from Hopkins et al. (2017) that an auditory cue interfered with an auditory task and that adding a third modality did not support overall performance in the dual-task, I decided to use visual cueing (i.e., with light, movement, or provided by a pedagogical agent) for the search tasks in the Experiments 2.1, 2.2, and 3 conducted for this dissertation. Thus, the search task and cueing method remained in the same modality (i.e., visual), whereas the learning task was presented auditorily, the aim

being to free up cognitive resources and reduce the conflict for resources, possibly also leading to superior learning outcomes. This was investigated in (desktop-) VR experiments, where the learning content was narrated during the search task, creating a dual-tasking VR learning scenario.

In addition to the cueing methods outlined above, it was also thought worthwhile to explore how pedagogical agents could provide cues and in what ways. Accordingly, the following section outlines the theoretical foundations regarding the use of pedagogical agents.

1.4.3 Pedagogical Agents

Pedagogical agents are defined as computer-generated humanlike virtual characters (Craig et al., 2002) that can be interactive and are designed to fulfil several functions (e.g., attention guidance and providing information) in learning materials (Moreno, 2012). Pedagogical agents could be ideally suited for VR learning environments, as VR allows a high degree of embodiment that seems desirable for pedagogical agents (Dai et al., 2022). Highly embodied agents show human-like features including facial expressions, synchronized lip movements while speaking, gestures, and subtle idle movements. By contrast, low-embodied agents tend to be more static or appear as voice-only agents entirely lacking visual representation (Davis et al., 2023).

In general, several theories, such as the social learning theory (Bandura, 1977), the social agency theory (Mayer, 2014; Mayer, Sobko, et al., 2003), and the cognitive-affective-social theory of learning in digital environments (CASTLE; Schneider et al., 2022), imply that adding social cues to learning material might be beneficial for learning. These theories add a complementary perspective to cognitive theories such as the CLT (Sweller et al., 1998, 2011) and the cognitive theory of multimedia learning (CTML; Mayer, 2014) that both emphasize learners' limited cognitive resources. According to CLT (Sweller et al., 1998, 2011) and the CTML (Mayer, 2014), it could be argued that including social cues either via a human on-screen instructor or a pedagogical agent without further didactic function could overload

participants. Moreover, pedagogical agents might also be perceived as “seductive details” – defined as interesting yet irrelevant parts of learning materials (Rey, 2012), which could also overload working memory and therefore be non-conducive to learning.

Specifically, CASTLE (Schneider et al., 2022) integrates several theories that emphasize the importance of social cues, such as the social learning theory (Bandura, 1977) and the social agency theory (Mayer, 2014), and states that including social cues in learning material activates learners’ social schemas, hence resulting in increased (para-)social processes. In turn, this activation affects the selection and further processing of information, which also influences the resulting schemas in the long-term memory (Schneider et al., 2022). According to CASTLE (Schneider et al., 2022), the inclusion of pedagogical agents could instil both verbal and visual social cues that enhance the social response (Lester et al., 1997). This enhanced social response is reflected in heightened brain activity in areas involved in social processes (Wenjing Li et al., 2022). Moreover, pedagogical agents may influence learning through factors such as motivation (Bian, 2022; Heidig & Clarebout, 2011; Sikström et al., 2022; Y. Wang et al., 2023) and improved perceived learning experiences (Grivokostopoulou et al., 2020).

However, several reviews (Martha & Santoso, 2019; Sikström et al., 2022) and meta-analyses (Castro-Alonso et al., 2021) on the impact of pedagogical agents on learning reveal only small positive effects on learning. Castro-Alonso et al. (2021) emphasise that the social behaviour (e.g., facial expression, gesturing, and motion) of the agents is often irrelevant and that these effects are mainly moderated by characteristics of the experiments (e.g., randomisation). By contrast, Dai et al. (2022) conclude that the design of the agents and in particular their degree of embodiment (e.g., expressed by gestures and facial expressions) is crucial and emphasize that a high degree of these characteristics tends to improve learning to a greater extent. Dai et al. (2022), however, point out that up until now only a small number of experiments have dealt with pedagogical agents in VR learning environments but that,

nevertheless, including agents in VR could enhance the learning experience and result in higher presence. Although the disparate conclusions of these studies – Castro-Alonso et al. (2021) and Dai et al. (2022) – can partly be attributed to the fact that the two analyses cover slightly different time periods (i.e., 2012–2019 vs. 2010–2021) and include a different number of studies (21 vs. 75) respectively, the disparity underscores the need for further research, especially utilizing newer technological advancements.

Since Experiment 3 in this dissertation focusses on attention guidance provided by pedagogical agents, the next section provides a general overview of the possible functions that pedagogical agents could fulfil.

1.4.4 Function of Pedagogical Agents

Pedagogical agents can fulfil different functions in learning materials, such as providing feedback, instruction, or guidance (Heidig & Clarebout, 2011), although acting as an information source is the most frequently implemented function according to Dai et al. (2022).

Particularly in learning material with undefined fields of view, attention guidance using position, gestures, or gaze shifts could be a central function for pedagogical agents in 360° environments. Researching into the potential of agents' positioning for directing attention, several studies with human on-screen instructors using eye-tracking technology suggest that humans (or human faces) attract attention in learning videos with a pre-defined field of view (Kizilcec et al., 2014; Sondermann et al., 2024). Moreover, since research from Speicher et al. (2019) implies that following a person in a 360° video is one of the most efficient methods for attention guidance, the assumption that the position of a pedagogical agent might also assist learners in finding the relevant field of view seems plausible.

Pedagogical agents can also use gestures, such as pointing, for attention guidance (Khokhar & Borst, 2022), which have been shown to positively affect learning (Davis, 2018; Wenjing Li et al., 2019). Furthermore, gaze movements from another human are an important

information source for a person's orientation (Bayliss et al., 2010; Chacón-Candia et al., 2023). This insight, along with the meta-analyses of McKay et al. (2021), suggests that gaze cueing may be equally effective regardless of whether it is provided by real or virtual faces. Thus, an agent's gaze could prompt joint attention, meaning that several people might have their attention directed to an object that may be of mutual interest (Frischen et al., 2007).

Some research has already shown that gaze shifts provided by a pedagogical agent successfully guide attention (Pejsa et al., 2015). Pejsa et al. (2015) also maintain that an agent's "referential" gaze (i.e., including head movement) at nearby learning material also improves learning outcomes. Contrary to this, the findings of Wenjing Li et al. (2023) suggest that the gaze of a pedagogical agent alone cannot successfully direct attention and does not improve learning. The depicted agent in this study also directed its head to an adjacent graphic. However, this agent did not further adjust its gaze to provide more detailed guidance regarding which part within the depicted graphic was the most relevant (for instance, by varying the height of its gaze). Thus, it probably only offered limited guidance in the gaze-only condition (only present in Experiment 2) (Wenjing Li et al., 2023). Therefore, whether a pedagogical agent's gaze combined with its position might successfully direct attention in a 360° environment remains an open research question.

Overall, while cueing may yield positive effects in various contexts, the following section explores various theories about unintended negative impacts cueing might have on the development of mental spatial representation.

1.4.5 Mental Representation

As advancing technologies increasingly simplify everyday tasks, there are growing concerns about their potentially detrimental effects on human cognitive abilities, particularly memory (for an overview, see Clinch et al., 2021). Moreover, research on "desirable" difficulties suggests that learning conditions that are perceived as challenging often result in deeper understanding (de Bruin et al., 2023). Therefore, not all support systems that initially

facilitate a learning task necessarily lead to superior outcomes. Given that cueing can be viewed as a form of digital assistance, it becomes crucial to explore whether it might negatively impact the development of mental spatial representation – the way we perceive, remember, and think about the spatial relationships between objects.

One area of daily life, in which people frequently seek digital assistance is navigation using GPS devices. However, recent research implies that relying on digital navigation devices leads to an insufficient development of mental spatial representation, resulting, for instance, in poor performance concerning navigation and landmark recall tasks (Dahmani & Bohbot, 2020; Gardony et al., 2015; Hejtmánek et al., 2018; Ishikawa et al., 2008). Ishikawa et al., for example, compared the spatial representation of participants in three groups in an unfamiliar city. The first group navigated with GPS, the second group used a map, and the third group was guided by an instructor (i.e., they first walked with an instructor and were later required to replicate the route on their own). Results showed that the GPS group took significantly less efficient routes and produced less precise maps during the testing phase. Possible explanations for these adverse effects are still being discussed. However, after conducting research including auditory navigation aids, Gardony et al. (2015) suggested that the divided attention caused by the navigational device played a decisive role. Thus, it seems reasonable to assume that visual GPS systems, which often run on a different device, might also avert attention from the environment. In line with this explanation, results from Stefanucci et al. (2022) imply that implementing visual navigation cues inside the users' field of view during a navigation task in VR enhanced spatial learning. However, it needs to be emphasized that the cues in the experiment from Stefanucci et al. (2022) provided additional information about distance.

Several explanations for these and other negative effects of digital support are still being discussed, for instance, the (un)intentional forgetting of digitally accessible information, which might be due to a lack of suitable encoding caused by a lack of practice (van der Haak,

2014). Further, cognitive offloading, which, as previously defined, can be understood as a process of externalizing cognitive processes (Grinschgl et al., 2023), might also influence memory performance. Even though cognitive offloading might be helpful in daily tasks, results from a recent review have implied adverse effects on learning (Skulmowski, 2023). Although supporting learners in a practice phase of a task by providing an offloading opportunity (i.e., allowing the manipulation of graphics to fit the task requirements) was shown to enhance immediate task performance, without the offloading opportunity, learners were not able to uphold their former performance level (Moritz et al., 2020).

Whether cueing that is provided inside the field of view without further information regarding an object's distance to the viewer influences the development of mental spatial representation will be investigated further in two of the upcoming experiments. However, in the next section, the theoretical background regarding further design considerations for VR learning environments is outlined.

1.5 Additional Design Considerations

Besides attention guidance, there might be additional design decisions that influence learning, such as the design of the pedagogical agent and the setting of the learning environment. Thus, the following sections will briefly outline theoretical considerations regarding design choices concerning pedagogical agents, especially focussing on their clothing, and considerations regarding the design of the setting.

1.5.1 Design and Clothing of Pedagogical Agents

Technical progress has enabled the creation of more human-like pedagogical agents. Even though a high degree of embodiment seems particularly promising (Dai et al., 2022), some features of pedagogical agent design have still been insufficiently researched.

The Pedagogical Agents Conditions of Use (PACU) framework, developed by Heidig & Clarebout (2011), outlines four key conditions for the effective use of pedagogical agents: (1) the learning environment in which pedagogical agents are to be implemented; (2) the

characteristics of the learners; (3) the functions of the pedagogical agent; and (4) the agent's design. In terms of design, the framework distinguishes between a global design level (e.g., human vs. non-human or static vs. animated), a medium design level (e.g., partial vs. full display, including audio output, and aspects of language style), and a detailed design level (e.g., the age, gender, and clothing of the pedagogical agent are included). However, although the design feature clothing is mentioned at the detailed design level, it has received little attention in research.

Considerations regarding why the clothing of a pedagogical agent might play a significant role can be found in literature concerning the effects of clothing for humans. In general, clothing conveys information about the person wearing it (Damhorst, 1990). Moreover, behaviour towards a person can also be influenced by clothing and the meaning attributed to it (Blumer, 1969). This especially applies to clothing that is typically associated with a particular professional group (e.g., police officers, doctors, judges, priests, soldiers, or craftsmen), and this association often leads to an attribution of competence (Gurney et al., 2017; Kodžoman, 2019) and trust (Hartmans, 2013). In the field of education, research has also shown that presentations given by formally dressed students were rated more favourably than those given by casually dressed students (Gurung et al., 2014). Further, formally dressed teachers were perceived as more knowledgeable and well prepared compared to their more informally dressed colleagues (Lukavsky et al., 1995). In addition, Beege et al. (2019), who focused on the interaction between clothing (professional vs. non-professional) and addressing style (frontal vs. lateral) in an educational video, found that the professionally dressed teacher improved learners' retention performance when teachers looked directly at the camera (Beege et al., 2019).

Therefore, it would appear important that pedagogical agents are also appropriately and professionally dressed. However, there has been little research on this topic. Nevertheless, Petersen et al. (2021) compared the thematical appropriateness of an agent in a museum

context (i.e., dressed as a museum guide vs. the same agent as a black mesh) and behavioural realism (high vs. low, respectively) with a control group without an agent and found heterogeneous results. Whereas most of the pedagogical agents hindered the acquisition of factual knowledge, the appropriately dressed agent with low realism (i.e., without gestures, limited idle animations, and without lip-syncing) had a positive effect on conceptual learning. However, in this research, while the agent's attire as a museum guide was suitably aligned with the surrounding environment (i.e., a museum), it did not specifically relate to the educational topic at hand, which was virology. Bearing in mind that Veletsianos (2010) recorded promising results on learning when the appearance of the agent matched the teaching content regarding facial features, testing whether these effects also occur when the clothing of an agent matches its teaching content seems also relevant.

To conclude, since research regarding the effects of a pedagogical agent's clothing is inconclusive, and to ensure the credibility of the agent that was designed to provide the cueing in Experiment 3, Experiment 1 was conducted. Moreover, because the interaction between the clothing of an agent and the setting in which the agent is positioned is yet another open research question, Experiment 1 also deals with the possible interactions between clothing and setting.

1.5.2 Setting

In educational contexts, the question arises whether an authentic setting, i.e., a setting that closely resembles a professional context, should be used and, if so, how it should be designed and implemented (Anker-Hansen & Andreé, 2019). One argument in favour of implementing authentic settings would be that several studies imply that more authentic environments, such as research facilities for scientific domains, have a positive influence on students' perception of relevance and interest (Betz, 2018; Schüttler et al., 2021). The extent to which these promising results from real-world settings can be transferred to digitally created learning materials such as videos or VR environments is still questionable. Moreover,

this question applies equally to photorealistic material (e.g., videos or 360° videos) and to digitally created material (e.g., digital video backgrounds or VR environments) (Catrysse et al., 2023; Merkt et al., 2020; Schroeder & Adesope, 2012). In the case of photorealistic material, the setting can be understood as the physical background in front of which a video was shot (see Merkt et al., 2020), while with digitally created material, the setting can be understood as the digitally designed background in which the virtual scene unfolds.

One possible effect of an authentic setting might its impact on the evaluation of the expertise attributed to the instructor, which in turn may affect learning (Merkt et al., 2020). Therefore, Merkt et al. (2020) compared a video about floral topics using either a botanical garden (i.e., an authentic setting) or a white wall (i.e., a neutral setting). The setting had no consistent effect on learning outcomes or perceived expertise in two experiments, leading the authors to the cautious conclusion that the effect of setting might be negligible. Similarly, Catrysse et al. (2023) investigated whether an authentic background compared to inappropriate and neutral backgrounds in an educational video about glaciers had an influence on learning outcomes. They employed a photograph of a geography classroom as an example of an authentic setting, a grey wall as a neutral setting, and an image of a beach club as an inappropriate setting. Additionally, they examined the potential influence of working memory capacity in this context. In line with the results of Merkt et al. (2020), Catrysse et al. (2023) found no consistent effect of setting, though their results suggest that unsuitable backgrounds are particularly detrimental to learning outcomes for students with low working memory capacity.

However, since the studies outlined above used either photorealistic backgrounds or photographs as authentic settings, the question of whether there might be setting effects on the assessment of expertise and subsequent learning for completely digitally created and animated learning videos remains. Moreover, the experiments mentioned above utilized human teachers. As such, exploring the effects of fully animated videos featuring pedagogical agents

as animated instructors could also be worth investigating. Pedagogical agents may employ fewer indications of expertise due to their artificial creation and may hence be more susceptible to other expertise cues, such as the virtual setting they are operating in or the cues provided by their clothing. Therefore, in Experiment 1, I investigated whether the setting of a digitally created learning video and the clothing of a pedagogical agent influenced learning and perceived expertise. Moreover, I aimed to ensure the thematic fit and credibility of the setting for all following experiments and the pedagogical agent used in Experiment 3.

2 Dissertation Overview

Based on the presented theoretical background, the following section outlines the research conducted in this dissertation, which includes four experiments distributed across three manuscripts: Experiment 1 is detailed in the first manuscript, Experiments 2.1 and 2.2 are combined in the second, and Experiment 3 is presented in the third manuscript.

Regarding Experiment 1, past research has indicated that the perceived thematical fitting of the setting (Catrysse et al., 2023; Merkt et al., 2020) as well as the perceived thematical appropriateness of a pedagogical agents' clothing (Petersen et al., 2021) might impact learning, perceived expertise, and credibility. Thus, Experiment 1 investigated the effects of these variables on the perceived professionalism of the video, the perceived expertise of the pedagogical agent, and learning using a digitally created learning video depicting a virtual workshop. Moreover, the experiment focused on ensuring that the setting and the design of the pedagogical agent were perceived as thematically appropriate and credible and can thus be used in the (desktop-) VR experiments.

With respect to Experiments 2.1, 2.2, and 3, the signaling principle (Alpizar et al., 2020; Schneider et al., 2018; van Gog, 2021) suggests that cueing positively influences learning outcomes and orientation across a wide range of learning formats, particularly when learners have a predefined field of view (de Koning et al., 2007, 2011; L. Lin & Atkinson, 2011). By contrast, research on the effects of cueing in learning environments with an undefined field of view (i.e., learners have to choose it themselves) is limited. Accordingly, two aspects of cueing are examined in this thesis: on the one hand, I investigated whether the guidance of attention to the search objects in a VR environment was successful, and, on the other hand, I also tested whether this orientation aid had a positive effect on learning outcomes. The search task with the simultaneously narrated learning content constituted a dual-task. Given the previous research on dual-tasks, it seemed reasonable to assume that by facilitating the search

task, the overall performance might improve due to a redistribution of the learners' cognitive resources (Grinschgl et al., 2023; Hornof et al., 2010).

Focussing on attention guidance, previous research showed promising results using movement and light cues (Rothe & Hußmann, 2018). Further, research on pedagogical agents also implied that these might be useful for attention guidance (Pejsa et al., 2015). Thus, for this dissertation, these methods were investigated in the VR search task. Experiment 2.1 and 2.2 tested a movement and light cue in desktop-VR and Experiment 3 tested a pedagogical agent and light cue in HMD-VR.

Moreover, several studies have implied that the use of navigational aids might impair spatial memory (Gardony et al., 2015), and research on desirable difficulties moreover has suggested that learning perceived as more difficult often deepened learning (de Bruin et al., 2023). Thus, I further investigated whether cueing impaired the formation of mental spatial representation in Experiments 2.1 and 3. In this context, the adverse effects of cueing may stem from learners becoming overly dependent on the provided cues (van der Haak, 2014), which could result in a subpar encoding of the spatial information.

Finally, because according to CAMIL (Makransky & Petersen, 2021), presence is a central prerequisite for learning in VR, I additionally investigated how different cueing methods influence presence in Experiments 2.2 and 3.

All the experiments were approved by the local ethics committee of the German Institute for Adult Education. Participants in all experiments gave informed consent prior to the experiments and were debriefed upon completion. For an overview of the experiments, including the links to the pre-registrations and data scripts, see Table 1. In the following section, I will outline the four experiments and their key variables, hypotheses, and results. For a more comprehensive description, please refer to the respective manuscripts in Appendices A, B, and C.

Table 1

Overview of the experiments

	Manuscript 1	Manuscript 2		Manuscript 3
	Experiment 1	Experiment 2.1	Experiment 2.2	Experiment 3
Setting	Online	Online	Online	Laboratory
Design	2x2 between-subject design	1x3 within-subject design	1x3 between-subject design	1x3 between-subject design
Independent variables	Clothing: (Thematically appropriate vs. inappropriate) Setting (Thematically appropriate vs. inappropriate)	Cueing: (Movement, light, none)	Cueing: (Movement, light, none)	Cueing: (Pedagogical agent, light, none)
Materials	Animated learning video	Desktop-VR-learning environment	Desktop-VR-learning environment	HMD-VR-learning environment
Dependent Variables	Learning	Learning Search times Mental spatial representation	Learning Search times	Learning Search times Mental spatial representation

			Presence Cognitive load	Presence Cognitive load
	Professionalism of the video Expertise of the agent Credibility of the agent Thematical fit of the clothing Thematical fit of the setting			
Sample	*N* = 200 (Prolific)	*N* = 60 (Prolific)	*N* = 159 (Prolific)	*N* = 91 (Institutional database)
Pre-registration	https://osf.io/bnv32?view_only=e19719a4eee243f8b40fb2e85add8d07	https://osf.io/vm6xj?view_only=6c07aaea38d249fdaeb35638a1d068b4	https://osf.io/9v4rk?view_only=c8a4e050839542f1ba3ecb10c32e1176	https://osf.io/7ufrh/?view_only=28ae330c9bb344b191b66428d1f50b83
Data and scripts	https://osf.io/qm4pv/?view_only=1715aaa02cb74e588b591882b08a3bd3	https://osf.io/2f87s/?view_only=173e32e228e84510b8f293a0dcd92898		https://osf.io/u6z8k/?view_only=dcdd029041624a8e954b1303909b38ff

2.1 Experiment 1

As outlined above, Experiment 1 aimed to ensure the credibility and thematic appropriateness of the pedagogical agent that was to be used in Experiment 3, as well as the credibility and thematic appropriateness of the settings for all the subsequent (desktop-) VR experiments. Thus, I created four versions of an animated learning video (i.e., a 2x2 design) that varied the clothing (thematically appropriate vs. thematically inappropriate) of the pedagogical agent and the setting (thematically appropriate vs. thematically inappropriate) of the video to investigate the effects of these variables on the agent's credibility and the perceived thematical fit of the clothing and setting to the learning content. Regarding these variables, I assumed that the thematically appropriately dressed agent would be perceived as more credible and would be evaluated as a better fit to the learning content than the inappropriately dressed agent. Moreover, I assumed that a thematically appropriate setting would also be evaluated as a better fit to the learning content [1].

Beyond these variables, which would be relevant for further studies, I also pre-registered additional hypotheses that explored the effects of clothing and setting on the perceived professionalism of the learning video, the perceived expertise of the pedagogical agent, and the learning outcomes. Considering the prior research outlined above, I hypothesized that the video showing the pedagogical agent wearing thematically appropriate clothing would be rated as the most professional. I moreover assumed that the thematically appropriately dressed agent would be considered to have more expertise and would achieve better learning outcomes. Since previous research revealed mixed results regarding the effects of setting on learning outcomes (see Merkt et al., 2020), I did not hypothesize a main effect of setting on learning.

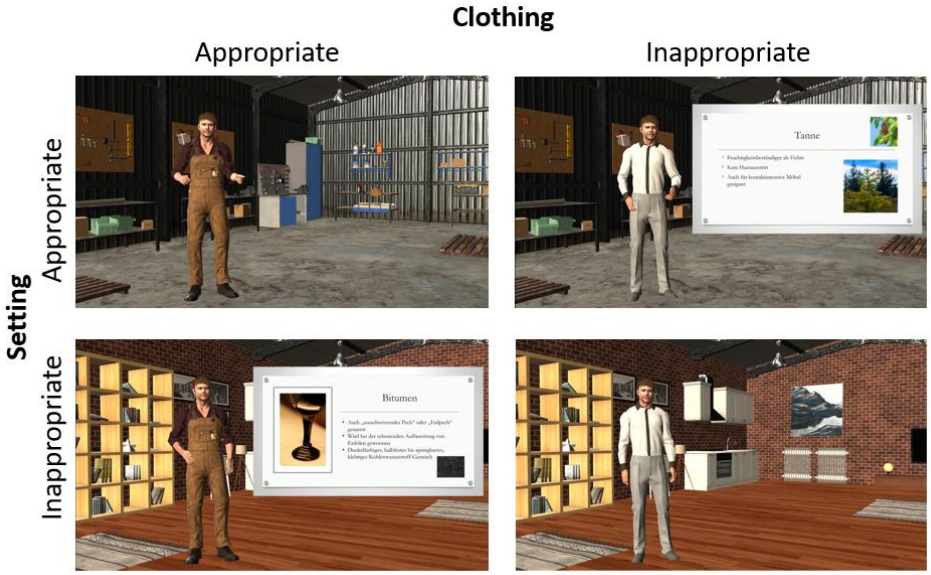
[1] In game development, assets are digital resources such as 3D models and textures used in game engines.

Overall, 231 participants, who were recruited via the online panel provider Prolific, completed the online experiment using LimeSurvey. Due to pre-registered exclusion criteria, 31 participants were excluded, resulting in a final sample of 200 German-speaking participants (131 male, 68 female, 1 diverse; M_{age}= 29.52; SD_{age}= 8.73). Participants were randomly assigned to one of the four experimental conditions. Regarding the experimental procedure, the experiment started with a picture and sound test, and then assessed prior knowledge and general interest in the content. Afterwards, the learning video representing the participants' assigned condition was presented (duration: 19'30 minutes). In this video, a pedagogical agent narrated learning material on craft content, which was partly visualized with a PowerPoint presentation to the right of the agent. The presentation only highlighted important content; hence, it was occasionally superimposed over the background, although the video's setting was visible for the remainder of the video. The pedagogical agent featured a full-body representation with a male human voice, displaying conversational gestures (see Davis, 2018) and mimicking but without further deictic functions. The experimental conditions resulting from the 2x2 design were visually implemented in the videos as follows: In the appropriate clothing conditions, the pedagogical agent wore brown dungarees, inspired by the recommendations for workwear of the Employer's Liability Insurance Association for Wood and Metal in Germany (BGHM, 2019). Regarding the inappropriate clothing condition, the pedagogical agent wore formal clothing (i.e., a white buttoned shirt with neat grey trousers). For the thematically appropriate setting, a prototype workshop with a concrete floor, a workbench, and garage-like walls were used, whereas the inappropriate condition resembled a living room. The number and distribution of assets in both settings were chosen to match as closely as possible. For an illustration, see Figure 2. The video was followed by evaluation questions regarding the credibility of the pedagogical agent, the thematic fit of the agent's clothing and the setting to the topics presented, the agent's expertise, the video's

professionalism, a 20-question multiple-choice knowledge test, and variables assessing exclusion criteria and demographic data.

Figure 2

Conditions of the learning video displaying the variation of setting and clothing



Note. To illustrate the insertion of the presentation, one version with and one without the presentation was selected for each setting.

The results revealed that the appropriately dressed pedagogical agent was rated as more credible and more fitting to the teaching content than the thematically inappropriately dressed agent. Likewise, the thematically appropriate setting was rated as more fitting to the content than the thematically inappropriate setting. However, there were no main effects or interactions concerning the perceived professionalism of the video and the perceived expertise of the agent. Moreover, there were no significant main effects on learning regarding either clothing or setting. However, there was a significant interaction regarding learning, where Bonferroni-adjusted post-hoc comparisons showed that in the appropriate clothing condition, the inappropriate setting resulted in better learning outcomes than the appropriate setting. Even though this interaction might point to a possible boundary condition, given the small

effect size (η_p^2 = .02), further research should be conducted before jumping to premature conclusions.

A possible explanation for the lack of systematic effects regarding the pedagogical agent's clothing and video setting on perceived expertise, professionalism, and learning might be that even in the inappropriate mode, they were still perceived as moderately fitting. Moreover, regarding learning, the appropriate setting might have created the unmet expectations of practical demonstrations, which, according to expectancy violations theory (Burgoon, 1993; Dunbar & Segrin, 2012), could have had detrimental effects on learning. However, practical demonstrations were not included because the four conditions needed to be kept constant.

To conclude, while designers may have some flexibility regarding clothing and setting, a highly inappropriate attire and setting could still affect learning outcomes. The limitations of this experiment, however, were the testing of only a single exemplar in both the appropriate and inappropriate clothing and settings. This highlights the need for future research to incorporate a broader range of assets across various disciplines. Future research could also contrast thematically appropriate clothing with attire strongly associated with other occupations. Moreover, regarding the suitability of the assets, the appropriate clothing of the pedagogical agent and the appropriate setting were all rated as appropriate for the craft teaching content. Thus, these assets were employed in the (desktop-)VR environment used in the following experiments.

2.2 Experiment 2.1

In this experiment, participants had to solve a search task while simultaneously listening to narrated learning content in a virtual workshop using the assets pre-tested in Experiment 1. In this learning environment, I investigated the effect of the independent variable cueing with the factor levels movement and light compared to a control condition without cueing on search times, learning, and mental spatial representation. Based on the theoretical framework

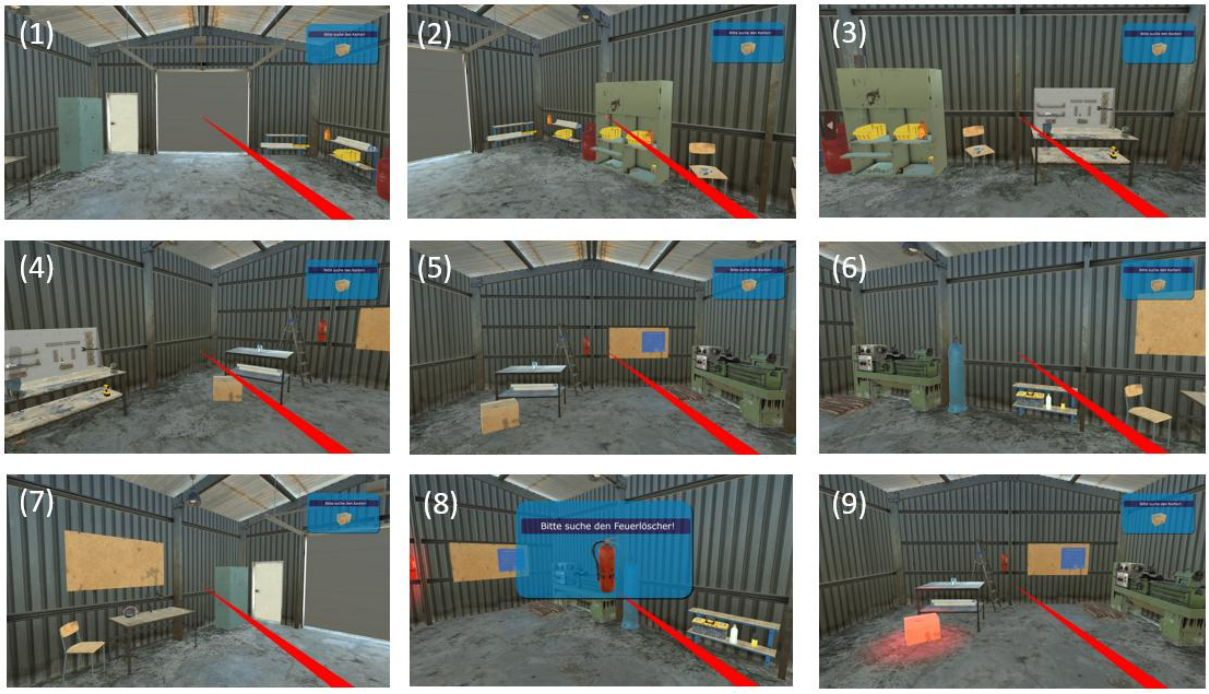
outlined above, I preregistered the hypotheses that light and movement cues will result in faster search times. Moreover, I assumed that the learning outcomes would be better when the search task was supported by cueing due to a redistribution of cognitive resources. Finally, I hypothesized that participants would form a more accurate mental spatial representation of object locations when solving the search task without cueing support.

Experiment 2.1 was completed by 70 participants using the online panel provider Prolific for recruitment and LimeSurvey for conducting the experiment. Ten of those participants had to be excluded due to the pre-defined exclusion criteria, resulting in a sample of 60 German-speaking participants (38 males, 22 females; $M_{\text{age}} = 29.00$; $SD_{age} = 7.62$). The independent variable cueing was varied within-subjects. The assignment of the cues to the search objects was balanced, and each object was represented with each cueing method across participants to avoid sequence effects. Participants were randomly assigned to one of the six resulting sequences. After this assignment, participants started downloading the corresponding version of the learning environment onto their PCs. Then, a picture and sound test followed, along with the evaluation of prior knowledge, interest in the learning content and VR experience. Finally, participants were instructed on how to run the desktop-VR learning environment. In the virtual workshop, participants had to locate 16 objects (15 + 1 trial object at the start) with a virtual laser pointer controlled with the PC mouse. The objects were evenly distributed in the learning environment. While participants searched for the various handicraft objects, such as hammers and brushes, a male human voice from off-screen narrated learning content about these objects (duration per object: between 27 and 75 seconds). The search tasks were shown on a user interface (UI) for 4 seconds, containing a picture of the object and a search prompt such as “Please find the hammer”. After this UI had disappeared, a smaller version remained in the upper right corner. When participants selected the correct object, a confirmation UI was superimposed over the learning environment until the narration for the object was finished. If participants were unable to find an object after 300

seconds, the task was registered as incomplete, and the next search task began. At the beginning of each task, participants' FOV was automatically returned to the same starting position within the 360° scene. The cueing was implemented as follows: in the movement cue condition, the cued object moved slightly in three-dimensional space ("shaking"), and in the light cue condition, the cued object was illuminated with a reddish light that imitated the light of a projector. For an illustration of the learning environment, see Figure 3.

Figure 3

The learners' perspective of the learning environment



Note. (1)–(7) Step-by-step sections of the learning environment in a clockwise direction; (8) Example of a UI with a search prompt; (9) Example of an object with a light cue.

While participants solved the search task, the learning environment recorded the search times (in seconds[2]). After completing the learning environment, the learning outcomes were

[2] This unit (i.e., seconds) was used because it ensured reliable measurements using the Unity Delta Time on computers with different performance levels that display the learning environment with different FPS numbers.

assessed using a knowledge test of 45 open-ended questions in a short answer format (3 for each of the 15 objects) dealing with factual knowledge from the narration. These open items were later rated by two independent raters assessing the correctness of the answers with a coding scheme. Then, the participants' mental spatial representation of the spatial structure of the VR environment was captured using 15 screenshots (i.e., one per object) of the environment. In these screenshots, the 15 search objects were not displayed. Instead, participants were asked which search object used to be in a marked location in the screenshot during the learning environment (see Figure 4 for an example). Finally, the demographic data were collected.

Figure 4

Mental spatial representation survey

Note. Participants were asked which item had been placed in the marked location (red square) during the search tasks.

The experiment revealed a significant effect on search times, with Bonferroni-adjusted post-hoc tests showing that search times were significantly reduced for objects highlighted with the movement cue compared to both other conditions, whereas no significant difference could be found between objects located with the light cue and objects located without cues.

Thus, the first hypothesis could only be confirmed for the movement cue. The light cue, however, did not facilitate searching, possibly because light cues highlight objects statically, attracting less attention than movement cues, which aligns with prior research indicating that movement attracts attention automatically (Jiang et al., 2018). Furthermore, in a within-subject design where participants were unaware of which cue would appear next, the static light cue, which guided attention in a less automatic way, may have been at a disadvantage.

Further, the experiment also revealed that cueing had no effect on both learning outcomes and on the formation of mental spatial representation, contradicting the second and third hypotheses. A possible explanation for the lack of effect on learning could be that the search times were drastically shorter than expected over all conditions (M = 10.27, SD = 5.67); hence, participants had ample time to solely focus on the narration, which took between 27 and 75 seconds (M = 44.13, SD = 13.30). Therefore, the overlap between the search task and narration was less substantial than intended, leading to a drastic shortening of the narration for Experiment 2.2. Regarding mental spatial representation, several factors may explain why it was not affected by cueing. One possible explanation was the generally low performance in the mental spatial representation test, potentially arising from a mismatch between the dynamic nature of the tasks and the static assessment methods employed. Another factor might have been the short search times and the accompanying covering of the environment after object selection. Thus, it is possible that cueing effects could have emerged if the mental spatial representation had been assessed in a more dynamic way and the learning environment had remained visible through the remaining narration.

To conclude, movement cues effectively guided attention, while light cues did not. However, neither cueing method impacted learning outcomes, possibly due to the limited overlap between the search tasks and learning content. The future experiments would thus shorten the narration to increase the overlap with the search task and additionally adopt a between-subject design to improve cue predictability.

2.3 Experiment 2.2

Employing the changes suggested by the results of Experiment 2.1 (i.e., using a between-subject design and a shortened narration), for this experiment, I preregistered the hypotheses that the search task would be solved faster if objects were highlighted by the movement cue. Moreover, I assumed that learning outcomes would be better when the search task in the learning environment was supported by the movement cue. To better understand the reasons for possible cueing effects, Experiment 2.2 also included the cognitive load scale by Andersen & Makransky (2021), which measures cognitive load on an 11-point Likert scale and was adapted to fit the learning environment[3]. This scale was specifically designed for immersive learning environments and considers extraneous cognitive load from interaction and environment, not just instructions. In immersive settings, interaction, environment, and instructions might each affect extraneous cognitive load and may hinder learning. Thus, differentiating these factors could be essential for optimizing instructional design in immersive media and for enabling meaningful comparisons with less immersive media. Thus, I further hypothesized that the environmental extraneous cognitive load (see the scale by Andersen & Makransky, 2021) would be lower in a movement cue condition. Since the light cue did not affect any of the dependent variables in Experiment 2.1, I exploratively investigated the effect of the light cue in this experiment. Further, I investigated the effect of cueing on presence, which was measured using a scale from Makransky et al. (2017) in the German version (Volkmann et al., 2018) that assesses physical presence and self-presence on a 5-point Likert scale.

Experiment 2.2 was completed by 179 participants recruited via Prolific, with the experiment running on LimeSurvey. Due to the predefined exclusion criteria, twenty participants had to be excluded. Thus, the final sample consisted of 159 German-speaking

[3] Originally the instrument covered laboratory safety, thus we adapted it to the virtual workshop in our experiments. Additionally, we used the term virtual environment instead of simulation.

participants (80 males, 77 females, 2 diverse; M_{age} = 30.23, SD_{age} = 9.11). Cueing was manipulated between subjects (factor levels: light, movement, and no cueing), and participants were randomly assigned to one of the three conditions.

In addition to the modifications resulting from Experiment 2.1, namely the implementation of a between-subject design and the shortened narration (i.e., 8 to 16 seconds per object; M = 12.91; SD = 2.34), in Experiment 2.2, participants only solved 12 search tasks (11 search objects and 1 practice trial at the beginning). Apart from these changes, the environment resembled the one used in Experiment 2.1. Furthermore, the procedure was similar to Experiment 2.1, with an additional assessment of cognitive load and presence after the learning environment but no assessment of mental spatial representation. Learning outcomes were measured using 22 open items selected from the item pool from Experiment 2.1. These items were chosen after removing those no longer covered by the shortened narration. From the remaining items, those that were too easy or too difficult were further excluded. The correctness of the answers was assessed by two independent raters according to a coding scheme.

The results revealed a significant effect of cueing on search times, with Bonferroni-adjustment post-hoc tests revealing that both light cue and movement cues led to faster search times compared to the condition without cueing. The movement and light cue conditions did not differ significantly. Thus, the first hypothesis that the movement cue reduces search times could be confirmed. Furthermore, explorative analyses also revealed an effect for the light cue, indicating that both cues effectively guided attention. Regarding the learning outcomes, this experiment revealed a significant effect of cueing, with Bonferroni adjusted post-hoc tests showing that the light cue led to better learning outcomes than the movement cue. The light cue, however, did not differ from the condition without cueing, and neither did the movement cue. Thus, the second hypothesis could not be confirmed. The finding that only light cues facilitated learning, while movement cues did not, may be elucidated by the results of the

exploratory analyses on presence. The explorative analyses revealed a significant effect of cueing on physical presence, with Bonferroni-adjusted post-hoc tests showing that the movement cue lowered physical presence compared to the condition without cueing. The movement cue, however, did not differ from the light cue, and the light condition did not differ from the condition without cues. Regarding self-presence, there was no effect of cueing. However, since CAMIL (Makransky & Petersen, 2021) emphasizes the importance of presence for learning in VR, the disruption in presence caused by the movement cue could have hindered learning.

Regarding the third hypothesis, this experiment revealed no effects of cueing on extraneous cognitive load on any of the subscales, implying that this hypothesis could not be confirmed. Overall, participants reported low cognitive load. The low cognitive load ratings could be attributed to the design of the dual-task, where participants may have focused primarily on the search tasks; hence, only considering this task in their rating (Hazeltine et al., 2002). If this was indeed the case, the short search times and low extraneous cognitive load ratings suggest that participants probably perceived the search task as relatively easy, which might also explain why cueing had no significant impact on their cognitive load.

Based on these results, for Experiment 3, I intended to test a cueing method that guided attention efficiently without reducing presence and to compare it to the light cues from the Experiments 2.1 and 2.2. My assumption was that as pedagogical agents could use their position, and gaze to guide attention, using them to implement cueing might be worthwhile (Pejsa et al., 2015; Speicher et al., 2019). A pedagogical agent might, moreover, add additional social presence to the learning material.

2.4 Experiment 3

In Experiment 3, I compared the effects of cueing provided by a pedagogical agent to the light cue and a condition without cueing in HMD-VR. Moreover, I intended to test whether the light cue worked as well in HMD-VR as it did in desktop-VR. For Experiment 3,

I preregistered the hypotheses that the search task would be solved faster if the object was highlighted by either a pedagogical agent cue or a light cue. Moreover, I assumed that the learning outcomes would be better when the search task in the learning environment was supported by either cueing method. I further assumed that the mental spatial representation of the learning environment would be better when the search task was not supported by cueing. Finally, I hypothesized that social presence would be highest when the search task was supported by the pedagogical agent cue.

This laboratory experiment was completed by 91 participants (46 males, 45 females; M_{age} = 37.44; SD_{age} = 14.28), none of whom had to be excluded. Cueing as the independent variable, with the factor levels pedagogical agent cue, light cue, and no cue, was manipulated between subjects. Participants were randomly assigned to one of these conditions. Regarding procedure, as in the previous desktop-VR experiments, prior knowledge, interest in the content, and VR experience were first assessed.

The VR learning environment in Experiment 3 was then presented on a Pico Neo 2 Eye, which is a stand-alone VR headset with six degrees of freedom (DoF) tracking, a 101° field of view, and a 75Hz refresh rate. The VR environment and tasks resembled the environment from Experiment 2.2 with only a few changes. Since in Experiment 3, participants had to physically turn around to perceive the whole scene and select the search objects with the right VR controller, participants were not automatically returned to a fixed starting position for each task to avoid motion sickness. Instead, they started the next task at the position where the last one ended, unlike in Experiment 2.2. To accommodate this change, some object positions were slightly altered to avoid the subsequent search object occupying the same FOV as the one in the previous task. After the UI with the prompt for the search task disappeared, a small thumbnail of the UI appeared on the virtual pinboard to enable participants to recall the current task. The identical narration from Experiment 2.2 was either presented by the pedagogical agent or via an off-screen voice. The selection of the correct object was

confirmed with a green check appearing in the middle of the participants' field of view, which remained there until the narration was completed.

The cueing was implemented by placing the pedagogical agent in close proximity to the search objects, using a full-body representation with conversational gestures that did not include deictic functions (Davis, 2018), and using speech mimicry while presenting the content. At the beginning of each task, the agent looked at the learners and then performed occasional gaze shifts towards the search objects, giving the learners additional hints about the objects' position. Before the start of each task, the agent was repositioned during a black overlay between tasks. The pedagogical agent was designed as an adult male and wore the craft clothing that was rated as credible in Experiment 1. The light cue was implemented as in Experiments 2.1 and 2.2. After finishing the learning environment, participants evaluated their learning experience regarding cognitive load and presence using the same instruments as in Experiment 2.2, with social presence being additionally measured in this experiment. Afterwards, participants completed the same knowledge test as in Experiment 2.2 and solved the mental spatial representation test that assessed how well the participants remembered the positions of the search objects. The mental spatial representation test resembled the learning environment and was presented on the PICO Neo, with two changes: red question marks were inserted as placeholders instead of the search objects; and there was no narration. The mental spatial representation test comprised twelve tasks mirroring the learning environment, each introduced by prompts such as "At which position was the hammer located?" Participants were instructed to identify the location of the object for each task by selecting the corresponding question mark that indicated its position. Participants had a maximum of 30 seconds per object and could within this timeframe repeatedly attempt to locate the correct placeholder. For an illustration, see Figure 5. Finally, participants provided their demographic data.

Figure 5

Mental Spatial Representation Test – Experiment 3

Note. Participants were asked which red question mark represented the search object in question in the learning environment and to select it using a virtual laser pointer.

Results revealed an effect of cueing on search times, while post-hoc tests with Bonferroni adjustment showed that the pedagogical agent and the light cue both led to faster search times compared to the condition without cues, confirming the first hypothesis. The pedagogical agent cue and the light cue were equally effective in guiding attention.

Regarding the second hypothesis, cueing had no significant effect on learning outcomes. A possible explanation for this could be that the participants did not perceive a lower extraneous cognitive load with cueing (i.e., there was no significant effect on cognitive load on any of the subscales). This suggests that the cueing did not free cognitive resources for deeper engagement with learning materials. Moreover, the active nature of the search task may have been the primary focus, reducing attention to learning materials and rendering cueing ineffective. In addition, the overall level of cognitive load in Experiment 3 was even

significantly lower than in Experiment 2.2, which further reinforces the likelihood of an even stronger focus on the search task.

The learning outcomes for the pedagogical agent could also possibly be closely tied to the results of presence that were analysed for the fourth hypothesis. Those results revealed that there was no effect of cueing on social presence. However, cueing significantly influenced physical presence and self-presence. For physical presence, a post-hoc test with Bonferroni adjustment revealed that the pedagogical agent cue significantly lowered physical presence compared to the condition without cueing. However, it did not significantly lower the physical presence compared to the light cue. Moreover, the light cue did not significantly differ from the condition without cueing. Regarding self-presence, a Bonferroni adjusted post-hoc showed that the pedagogical agent cue decreased self-presence compared to the other conditions, whereas the light cue did not differ from the condition without cueing. Thus, the fourth hypothesis could not be confirmed.

Possible explanations as to why the pedagogical agent did not influence social presence might have had something to do with the agent's voice, which was the same in all conditions. Although it was off-screen in the conditions without the agent, it may have provided a consistent level of social presence. Moreover, the short duration of the learning environment possibly hindered the establishment of a social connection. Further, the decline in physical and self-presence may have been due to the agent's lack of representational fidelity, despite attempts at realism, potentially causing an “uncanny valley” effect (Mori et al., 2012). This effect suggests that as the resemblance to humans increases, the viewer's emotional response becomes more positive, but when the resemblance is close yet imperfect, it suddenly dips into a "valley" of unease or revulsion before rising again as the likeness becomes fully human. As in Experiment 2.2, this reduction in presence might have hindered more successful learning in the pedagogical agent condition (Makransky & Petersen, 2021).

Regarding the third hypothesis, results revealed no effect of cueing on mental spatial representation. Thus, the third hypothesis could not be confirmed. However, the absence of adverse effects of cueing on participants' mental spatial representation aligns with previous research, suggesting that cueing within the field of view does not have the same negative impact as navigational aids on separate devices, which often result in divided attention (Stefanucci et al., 2022). Maintaining a consistent perspective between the learning environment and the mental spatial representation test may have additionally facilitated the overall good performance in the mental spatial representation test, preventing the negative effects of cueing.

3 Discussion

In this dissertation, I conducted four experiments. Experiment 1 focused on ensuring the thematic appropriateness of the assets for the setting in all upcoming (desktop-) VR environments, as well as the agents' clothing for Experiment 3. The three subsequent experiments investigated the effects of different cueing methods (light cues, movement cues and cueing provided by a pedagogical agent) on search times and learning outcomes in desktop-VR and HMD-VR environments. Additionally, Experiments 2.1 and 3 measured the mental spatial representation of the layout, while Experiments 2.2 and 3 measured presence and cognitive load. Experiments 2.1 and 2.2 were completed online and used desktop-VR, while Experiment 3 was conducted in a laboratory and used HMD-VR.

3.1 Cueing in (desktop)-VR: Search Times

In total, three of the experiments investigated the effects of different cueing methods (i.e., movement, light, and pedagogical agent) in desktop-VR and HMD-VR environments on search times. The results imply that all cueing methods could successfully reduce search times in a between-subject design. However, only the movement cue reduced search times in a within-subject design (Experiment 2.1). It would thus appear that light cues attract less attention than movement cues because they are more static, which has been corroborated in previous research (Jiang et al., 2018). Furthermore, given that in the within-subject design participants had to deal with different cues from task to task, the accompanying lack of predictability may also have been disadvantageous for cueing methods that are less salient due to their static nature.

Addressing the issue of predictability, Experiment 2.2 investigated whether light and movement cues affect participants' search times in a between-subject design employing desktop-VR. In Experiment 2.2, both cues significantly decreased search time, hence both cues seem to guide attention successfully, supporting the assumption that while movement cues guided attention automatically in both predictable and unpredictable scenarios, light cues

need a predictable scenario in which participants can anticipate the upcoming cue to work effectively. Further, in a predictable design in HMD-VR, the light cue seemed to reduce search times just as effectively as in a predictable desktop-VR environment. The pedagogical agent cue also effectively improved search times. However, in this dissertation, evidence for this was limited to results obtained within a predictable HMD-VR environment. Furthermore, using a pedagogical agent cue in a within-subject design may introduce additional challenges other than predictability in that repeated appearance and disappearance of the agent across tasks could significantly undermine representational fidelity, potentially diminishing presence (Makransky & Petersen, 2021). Thus, testing how a pedagogical agent cue would perform in an unpredictable scenario is still an open research question. The following section discusses the extent to which these effects influence learning outcomes.

3.2 Cueing in (desktop-) VR: Learning Outcomes

Even though cueing seemed to effectively guide attention in most scenarios, it exclusively supported learning in a predictable between-subject desktop-VR scenario. The hypothesis that cueing in a dual-task might support overall performance was based on previous research that showed that supporting one task in a dual-task enhanced overall performance, for instance by training or cognitive offloading (A. D. Bender et al., 2017; Grinschgl et al., 2021). Since in Experiment 2.1 the search times were drastically shorter than expected the overlap between the search task and the narrated content was not as substantial as intended. Thus, the hypothesis that was based on the competition for cognitive resources might not have applied: participants had ample time to focus on the learning task, which might have rendered cueing ineffective.

In the experiments (2.2 & 3), in which participants actually had to engage in a dual-task, both experiments used a between-subject design to ensure the predictability of the upcoming cues. However, the light cue only positively affected learning in desktop-VR and resulted in significantly higher learning outcomes in desktop-VR (M = 8.76; SD = 3.50) than in HMD-

VR (M = 6.68; SD = 3.38). However, both these results were drawn from two different experiments using different mediums (i.e., one sample was conducted online and the other in the laboratory). Additionally, beyond employing a different medium, other disparities existed between these experiments. For instance, the mini user interface, which was positioned in the upper right corner of the participants' FOV in the desktop-VR, was placed on a virtual pinboard in the HMD-VR. Moreover, in the desktop-VR, participants began each task from the same starting position, while in the HMD-VR, they commenced the next task from the position where the previous one ended, to prevent motion sickness from a forced return to a fixed starting point. Despite these differences, this observation may also suggest that immersive HMD-VR may have presented greater challenges due to the physical movement required to perceive the environment and the greater extent of visual input, compared to navigating desktop-VR on a PC monitor using a mouse and keyboard. Additionally, the use of a headset in HMD-VR may have been more unfamiliar for many of the participants than the more conventional PC setup, which might also explain the poorer learning outcomes. To conclude, further systematic investigation is needed to understand why cueing failed to sufficiently alleviate cognitive demands and promote learning success in the more immersive HMD-VR, despite effectively aiding orientation. Nevertheless, the finding aligns with prior research by Makransky et al. (2019) that also reported decreased learning in immersive HMD-VR compared to desktop-VR, although this research did not focus on the effects of cueing.

One possible explanation for the difference in learning effectivity between desktop-VR and HMD-VR concerns cognitive load, which was not affected by cueing in either of the experiments. Thus, despite all cueing methods effectively reducing search times, participants did not seem to perceive their extraneous cognitive load as lower in the cueing conditions. Hence, cueing does not seem to result in a noticeable liberation of cognitive resources. In general, participants reported experiencing relatively low levels of all types of cognitive load in both experiments. However, it is noteworthy that participants rated their cognitive load for

every sub-scale significantly lower in HMD-VR than in desktop-VR, which seems counterintuitive but might be due to the tasks' design. Even though participants were informed about a knowledge test related to the narrated content in the instructions prior to the learning environment, the prompts within the VR environments focused exclusively on the search tasks. Additionally, the more active nature of the search task – reading the user interface, actively navigating the environment while scanning for the search objects – may have led participants to perceive the search tasks as the main priority. Therefore, the participants' attention might have been more captured by the active tasks, as the learning content only required passive listening. This might have led participants to base their rating entirely on the search task, which may have been considered easy, resulting in the lower ratings for extraneous cognitive load. The differences, especially in extraneous cognitive load, however, imply that the focus on the search tasks may have been so intense in HMD-VR that little attention was paid to the narration, rendering cueing ineffective. In general, perceiving one task as the main task is a frequently discussed issue in research on dual-task paradigms (Hazeltine et al., 2002). This interpretation also aligns with two studies (Jackson et al., 2023a, 2023b) that investigated dual-task performance involving a search task and a word recall task (i.e., a visual and an auditory tasks, respectively). Both studies found significant performance decrements in dual-task conditions compared to single-task performance. In the visual search task, participants navigated a simulated environment, searching for hidden objects (suitcases) while either receiving written or narrated words to memorize. This setup revealed that participants found significantly fewer search objects and recalled fewer words in the dual-task condition. Both studies highlight the issue of task prioritization, emphasizing that participants tend to prioritize the search task (i.e., the more active task) over the memory task (Jackson et al., 2023a, 2023b).

Another possible explanation involving cognitive load might also be related to the type of knowledge that should be learned. The narration mainly focussed on factual knowledge,

which according to Bloom's taxonomy is only the first level of processing since a deeper level of understanding is not necessary to recall information (Bloom, 1956). However, simply memorizing factual knowledge may not have been particularly engaging for participants, resulting in low germane cognitive load. However, given the challenges associated with dual-tasking, creating more difficult learning materials for these experiments did not seem prudent.

In line with that, the impact of cueing on presence should be considered. Presence was significantly higher in HMD-VR than in desktop-VR, as was to be expected given prior research (Shu et al., 2019; Slater & Sanchez-Vives, 2016). Considering the implications formulated in CAMIL (Makransky & Petersen, 2021), the HMD-VR should have therefore led to superior learning compared to the desktop-VR. However, since CAMIL (Makransky & Petersen, 2021) also implies that highly immersive HMD-VR environments might be particularly beneficial for procedural knowledge rather than factual knowledge, the higher presence in Experiment 3 may not have resulted in better learning due to a mismatch between the highly immersive learning environment and the type of knowledge conveyed in this experiment.

Nevertheless, it seems noteworthy that in both experiments the conditions that lowered presence, namely the movement and pedagogical agent cues, resulted in the poorest learning results, even though the effect on learning was only significant in Experiment 2.2. The reasons why the movement cues and the pedagogical agent cues lowered physical presence, and the pedagogical agent moreover reduced self-presence, may be attributed to their effect on representational fidelity. According to CAMIL (Makransky & Petersen, 2021), representational fidelity, which can be understood as the degree to which the representation accurately reflects reality (Dalgarno & Lee, 2010; Makransky & Petersen, 2021), is a key factor that influences presence. Thus, cues that lower representational fidelity might not be beneficial for learning. The movement cues may have disrupted presence because, in reality, objects do not typically move on their own. Regarding the pedagogical agent, the decline in

physical and self-presence could have been due to various factors, such as the agent's repositioning. This repositioning meant that at the start of each task, the pedagogical agent was repositioned to represent the new FOV in which the search object had been placed. Even though this repositioning was not directly visible to participants due to a black overlay, it could have been disruptive. Further, following recommendations by Dai et al. (2022), the goal for Experiment 3 was to create a highly embodied pedagogical agent with realistic movements such as breathing, gesturing, idle positions (i.e., slightly moving after the narration had finished), and facial expressions while narrating. While this approach aimed to design an agent that closely resembled a real person, it may not have achieved sufficient realism, potentially resulting in the so-called uncanny valley effect (Mori et al., 2012). This phenomenon suggests that the likability of artificial characters increases as they become more human-like but abruptly decreases when they become close to the level of real people. These characters almost resemble real people but still show slight hints of artificiality, which causes discomfort for many people. In light of this uncanny valley effect (Mori et al., 2012), creating fully human-like characters or clearly artificial ones could both be potential solutions. However, recent findings suggest that highly realistic video instructors might not enhance learning, possibly because they draw too much attention to themselves (Wenjing Li et al., 2024). Thus, it remains unclear whether the likely increase in the presence of more realistic, human-like characters, which may have the potential to influence learning, ultimately leads to improved outcomes. Overall, there may have been several reasons why presence was diminished by the movement and the pedagogical agent cues, whereas artificially illuminating objects may be more acceptable or “normal”, thus explaining why it did not reduce presence.

Furthermore, regarding social presence, the pedagogical agent unexpectedly did not have a positive effect. One possible explanation could be that the narration, delivered by the pedagogical agent, originated off-screen in the light cue and control condition, possibly conveying a level of social presence that could not be enhanced by the agent's visual image

(also see Mayer, Dow, et al., 2003). Additionally, the learning environment only lasted about five minutes, making it hard for participants to form a social connection with the agent. Moreover, the study took place in a laboratory setting where one member of the research team (and sometimes another participant) was present. This consistent social presence might have overshadowed the potential effects of the agent.

To conclude, these results shed light on the conditions under which cueing might be effective in a VR dual-task. Regarding attention guidance, the predictability of the upcoming cue seems to be crucial, as well as maintaining a high level of presence. However, the higher levels of presence that HMD-VR provided did not necessarily lead to better learning. Hence, investigating different learning content, such as procedural knowledge, may be worthwhile, as highly immersive HMD-VR may not be ideal for factual knowledge content. The following section discusses the results regarding mental spatial representation.

3.3 Cueing in (desktop-) VR: Mental Spatial Representation

Comparing the results from Experiments 2.1 and 3, cueing had no effect on mental spatial representation. However, the reasons for these results may differ between the experiments. Although in Experiment 3 participants picked the correct placeholder at the first try more often than they identified the correct missing object in Experiment 2.1, comparing these scores might offer only a limited insight due to the different measurement procedures.

Concerning differences regarding the learning environment, in Experiment 2.1 the learning environment was covered after the correct object was selected, whereas in Experiment 3 a green check mark was visible in the middle of the FOV, leaving the remainder of the environment still visible. This may have caused the overall lower level of mental spatial representation in Experiment 2.1. Furthermore, given the short overall search times, participants in Experiment 2.1 may not have had enough time to encode object locations. Moreover, in the HMD-VR environment, the participants perceived the objects in a “real”

physical space, which may have made it easier to remember the location of objects compared to the PC environment.

Further, in Experiment 2.1, participants were confronted with five search tasks representing each cue (plus the practice trial). Therefore, participants solved five search tasks without cueing. Thus, it cannot be ruled out that participants encoded other object locations during the tasks without cueing. As such, the between-subject design of Experiment 3 might be more suitable to investigate whether cueing influences the formation of a mental spatial representation. Moreover, in Experiment 2.1 several large objects were included (e.g., a cupboard and a stationary saw), which were excluded from Experiment 3. The reason for explicitly reducing the number of large objects in Experiment 3 concerned the differences regarding the learning environments and the assessment method. In Experiment 2.1, it may have been possible during the test phase to easily identify a missing object based on its size and the size of its corresponding mark in the screenshot, even without having encoded the object's position during the learning phase. Thus, the number of positions that participants genuinely remembered may have been even lower in this experiment. By contrast, Experiment 3 used question marks in the test phase that were the same size regardless of the size of the object they represented.

Regarding the assessment methods used in these experiments, there were also several differences. Experiment 2.1 assessed mental spatial representation with a static method (i.e., screenshots), which could have been detrimental since the tasks in the learning environment were dynamic. This resulted in a mismatch between the encoding and the testing. Moreover, learners to some extent may have also relied on spatial dynamic information (i.e., how far they had turned) to remember the object positions. Therefore, Experiment 3, which was conducted in the laboratory, employed a dynamic assessment method using the HMD.

However, the dynamic assessment method that maintained a consistent perspective with the learning environment and allowed the same movement in the 360° scene may have

facilitated the task because participants possibly also recalled their movements (e.g., that they had turned their head up or down for a specific object). Furthermore, in Experiment 2.1, participants were asked to name which object was in a marked position (i.e., “Which item was placed in the marked location?”). They therefore had to recall both the object’s name and associate it with a specific position. In Experiment 3, however, participants only had to specify the position of a given object (e.g., “At which position was the hammer located?'). Thus, in Experiment 3, to achieve a high score in the assessment, the participants only had to remember half the amount of information to achieve the same score as the participants of Experiment 2.1. Nevertheless, despite being easier, the assessment method in Experiment 3, which allowed participants to retry after selecting the wrong placeholder, provided broader insights as it not only evaluated whether participants succeeded but also when, and after how many mistakes, they achieved success. Overall, even though these differences mean that a direct comparison of results may not be ideal, the fact that the results from Experiment 3 were descriptively better than those in Experiment 2.1 is in line with the findings of Renganayagalu et al. (2021), which suggests that immersive media may be more advantageous for promoting mental spatial memory.

However, despite these differences, it is noteworthy that both experiments did not reveal an effect of cueing. Since Experiment 2.1 resulted in very low scores, the insignificant effect of cueing might have been due to floor effects. However, since in Experiment 3 with better scores there was also no effect of cueing, other reasons need to be considered. As research regarding desirable difficulties suggests that more perceived effort in the learning phase often results in better learning (de Bruin et al., 2023), the stronger retention of objects, evidenced by a low number of mistakes made, and the corresponding low germane cognitive load might indicate that participants did not perceive the task as difficult, regardless of cueing.

Further, as Gardony et al. (2013) suggest that impairments observed in previous studies on navigation aids may stem from a shift between first-person perspective navigation during

the learning phase and a survey-like assessment (e.g., viewing from above) during the test phase, maintaining the same perspective might alleviate some of those impairments. Moreover, the findings of both experiments align with research by Stefanucci et al. (2022), indicating that integrated cues provided within the environment may not have the same negative impact as navigational aids presented on separate devices like GPS tools. Integrated cues do not require users to divert their attention to an external device that could distract attention from the physical surroundings.

Furthermore, when participants are aware that their spatial memory will be tested later (as they were), they may place greater importance on the information. According to the cognitive architecture of digital externalisation outlined by Skulmowski (2023), a high perceived value of information leads to less offloading behaviour and greater engagement in memorizing the information. Conversely, a high extraneous cognitive load increases offloading. The results from Experiment 3 align with Skulmowski (2023), as participants, having been informed about the test and rating their extraneous cognitive load as low, memorized the object locations regardless of cueing. In summary, while no adverse effects of cueing were observed in these experiments, such effects cannot be entirely ruled out, particularly in scenarios involving a greater number of objects to memorize or larger VR environments. Finally, in the next section, I will discuss the limitations of the research conducted in this dissertation.

3.4 Limitations and Directions for Future Research

This research was subject to some limitations. Exploring alternative cueing methods in VR seems necessary because of the various advantages and drawbacks associated with each approach. Future research should concentrate on creating cues that integrate the strengths of the examined methods while avoiding their potential drawbacks.

First, the movement cue, as implemented in Experiments 2.1 and 2.2, might be applicable when the focus is on fast orientation instead of learning. However, whether

adaptations of the movement cue, such as using naturally occurring movements (for example, leaves moved by the wind or a ticking clock) might result in a different evaluation of presence and thus potentially better learning results could also be a valid path of enquiry. Further, the absence of measurements for presence and cognitive load in Experiment 2.1, due to concerns that self-assessment after each search task in a within subject design could have caused disruption, means other potential explanations (besides the limited task overlap) for why cueing did not facilitate learning in Experiment 2.1 cannot be excluded. This limitation in Experiment 2.1 may also highlight the need to explore alternative methods for measuring cognitive load in VR, such as utilizing eye-tracking to assess pupil diameter, as it might be possible to apply these measurements in real-time during the task (Souchet et al., 2022). However, since these measurements may leave room for interpretation, particularly regarding the distinction between visual fatigue and cognitive load, they may not entirely replace self-assessments (Souchet et al., 2022).

Further, the light cue could have limitations in scenarios with several objects in close proximity to each other, possibly emphasizing the need to investigate adaptive variations. These might include adjusting the broadness of the light cone or turning up the objects' inherent colour luminosity when the learners are far away and adjusting it back to a regular level to uphold representational fidelity when the learners get closer. Furthermore, the use of a flickering light could pave the way for intriguing research as it has the potential to combine the benefits of dynamic cues that naturally attract attention through movement and adaptiveness. This could be achieved by modulating the flicker frequency, for instance, accelerating the flicker to draw attention when participants are far away and reducing or halting it as they come closer to their target. Whether these adaptations would lead to a more automatic directing of attention and to what extent these methods could influence presence would be an interesting approach for further investigation, although these methods may also

influence representational fidelity as light does not usually flicker systematically in the real world.

The pedagogical agent cue was helpful to guide learners to a relevant field of view. However, subsequent gaze-cueing might also have limitations when several objects are in small areas because it might lack precision. Furthermore, in Experiment 3, only one pattern of gaze movements (i.e., looking at the participant, then looking at the object, and finally looking back at the participant) was displayed. Since this varied regarding the degree of head and upper body movement to reflect the search objects' position (i.e., the agent looked at certain objects by turning his upper body, whereas other objects were focussed using only head/eye- movements), future research should systematically investigate the effects of a pedagogical agent's gaze in a 360° scene. Moreover, as previously mentioned, relocating the agent between tasks may interrupt presence. Therefore, alternative methods could involve implementing smoother transitions between search tasks. For instance, a more seamless repositioning when the agent walks to a new location could mitigate the adverse impact on presence. Building on research by Speicher et al. (2019), which proposes that tracking a person's movements can effectively direct attention in VR videos, it would also be plausible to consider whether tracking an agent could similarly guide attention in VR learning environments. Moreover, since I only tested the pedagogical agent cue in a predictable between-subjects design, it remains an open question whether this method is capable of guiding attention in an unpredictable scenario. However, using a pedagogical agent in a within-subject design might negatively impact presence, as the disappearance and re-appearance between tasks might negatively affect representational fidelity.

As previously mentioned, there may be more than one possibility of improving cueing with pedagogical agents, as designing a "fully" human-like agent and an obviously cartoon-like depiction might both be valid choices to avoid the uncanny valley effect (Mori et al., 2012). In addition, the advantages of pedagogical agent cueing in applications over an

extended period should also be considered, as this might possibly encourage learners to feel a "social connection" with the agent.

Further, shortening the narration in Experiments 2.2 and 3 intentionally increased the overlap between the two tasks, possibly increasing overall task difficulty. However, since Experiment 2.1 did not include a measurement for cognitive load, this increased difficulty can only be assumed. Moreover, contrary to original expectations, this increased task difficulty was not reflected by high extraneous cognitive loads in Experiments 2.2 and 3 and, thus, did not necessarily lead to better learning outcomes when the difficulty of the search tasks was allegedly alleviated by cues, the exception being in desktop-VR with undiminished presence. Thus, systematically varying and measuring task difficulty in both desktop-VR and HMD-VR might allow further insights into possible boundary conditions of cueing in VR dual-tasks.

In the same vein, exploring the effects of cueing in more complex environments (e.g., larger virtual sets with more distractors) (Botch et al., 2023) could bolster understanding of cueing effects in different contexts. The findings of the current experiments may not necessarily apply to more complex VR environments because the used environment exclusively tested cueing in a medium-sized room where all objects could either be seen from the participants' position (Experiments 2.1 and 2.2) or could be seen with only a few steps in each direction (Experiment 3). Moreover, in these experiments, participants were aware that they needed to actively search for objects, prompting them to shift their field of view to locate them. Therefore, it would be intriguing to test these cueing methods in scenarios where the objects to be found, or the specific field of view that needed to be focused on, were not explicitly indicated, the only instruction being to explore the virtual environment. In this scenario, it would be intriguing to investigate whether cueing facilitates the most thorough exploration. This could provide further insights into the efficacy of cueing in more ambiguous contexts.

Moreover, as previously mentioned, the potential perception of one task as the primary task (probably the search task) may have influenced cueing effects on learning outcomes. This suggests the necessity for more dynamic and interactive learning task designs in VR learning environments. For instance, incorporating the manipulation of 3D models could significantly enhance active learning. The nature of these manipulations could vary greatly depending on the learning context. In a chemistry lab, participants might observe chemical reactions using 3D models of molecules, providing a tangible understanding of molecular interactions (Fombona-Pascual et al., 2022). In an engineering scenario, learners could engage in assembling machinery, honing their technical skills through a virtual yet realistic environment (Dolsak & Reich, 2022). Other examples might include surgical training where medical professionals practice complex procedures in VR (Gazit et al., 2024; Landau et al., 2024).

Furthermore, future research could delve deeper into the effects of spatial abilities and how they influence navigation in virtual environments to determine which cueing approaches are optimal for learners with varying spatial abilities. Recent research (Brucker et al., 2024; Wenhao Li et al., 2024) highlights the significant impact of spatial abilities on the effectiveness of learning comparing HMD-VR with desktop-VR. Specifically, Wenhao Li et al. (2024) found that learners with low spatial abilities demonstrated superior spatial knowledge acquisition when using HMD-VR compared to desktop-VR. By contrast, this disparity was not observed among learners with high spatial abilities (Wenhao Li et al., 2024). Thus, given that the results concerning mental spatial representation in Experiments 2.1 and 3 were not really comparable due to the different assessment methods, it could be worthwhile to further investigate how cueing might influence learners with varying spatial abilities in both types of VR regarding learning and the formation of a mental spatial representation. Moreover, the meta-analysis from Richter et al. (2016) suggests that prior knowledge is a boundary condition moderating the effects of cueing on learning in non-immersive learning

material. Hence, testing whether prior knowledge might also influence the effectiveness of cueing in VR learning environments could also be worthwhile.

3.5 Conclusion

To fully exploit the potential of VR learning environments, difficulties like user disorientation need to be addressed. Thus, three experiments investigated the effects of different cueing methods on search times (used as an indicator for successful orientation) and learning outcomes. Additionally, Experiments 2.1 and 3 measured mental spatial representation, while Experiments 2.2 and 3 assessed cognitive load and presence. Experiments 2.1 (within-subject design) and 2.2 (between-subject design) employed desktop-VR, whereas Experiment 3 used HMD-VR (between-subject design).

First, regarding search times, all cueing methods (movement, light, and pedagogical agent) were capable of reducing search times significantly in a between-subject design, whereas only the movement cue reduced search times in the within-subject design, indicating that movement guides attention more automatically, whereas light was more dependent on cue predictability.

However, only the light cue had a positive effect on learning outcomes in Experiment 2.2. Thus, in Experiment 2.2, the facilitation of search times seems to have freed up cognitive resources that led to better learning, even though there was no direct effect on self-reported cognitive load. With the movement cue, however, the facilitated search times did not increase learning. This may have been the result of decreased physical presence. However, further research is necessary to substantiate this finding. The pedagogical agent and the light cue both successfully reduced search times in HMD-VR. However, neither effect increased learning outcomes, possibly because in HMD-VR the focus on the search task was even stronger than in desktop-VR. Moreover, the pedagogical agent reduced self- and physical presence.

Overall, the advantages and drawbacks of the three cue types examined in these experiments must be carefully evaluated. The movement cue improved search speed but didn't

enhance learning, likely due to reduced presence. It may be best suited for tasks requiring quick navigation rather than learning. The light cue maintained presence but improved learning only in predictable desktop-VR scenarios, and its precision may be limited in crowded environments. The pedagogical agent effectively directed attention to the relevant field of view but reduced presence and did not improve learning, suggesting it may be less efficient than the light cue, particularly in scenarios where it is not necessary to demonstrate tasks. Moreover, designing a highly embodied pedagogical agent without triggering the uncanny valley effect (Mori et al., 2012) could be a significant challenge.

In conclusion, all the methods tested in this study were capable of directing learners' attention in predictable scenarios. However, only the light cue was helpful in a simultaneous learning task and only in a between-subject desktop-VR scenario – not in HMD-VR. Overall, regarding learning outcomes, there is room for improvement. Thus, dual-tasking in VR might not be a good choice for fact-learning situations, although it remains an open research question whether a dual-task is suitable for procedural learning. None of the tested cues negatively influenced the mental spatial representation of the environment. Future research should focus on developing cueing methods that dynamically guide attention without diminishing presence, hence upholding a high level of representational fidelity. These methods should be systematically tested in different virtual environments with varying task complexity to identify their effectiveness and explore potential boundary conditions.

4 References

Albus, P., & Seufert, T. (2022). Signaling in 360° Desktop Virtual Reality Influences Learning Outcome and Cognitive Load. *Frontiers in Education*, *7*, 1–10. https://doi.org/10.3389/feduc.2022.916105

Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education*, *166*, 104154. https://doi.org/10.1016/j.compedu.2021.104154

Alpizar, D., Adesope, O. O., & Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educational Technology Research and Development*, *68*(5), 2095–2119. https://doi.org/10.1007/s11423-020-09748-7

Andersen, M. S., & Makransky, G. (2021). The validation and further development of a multidimensional cognitive load scale for virtual environments. *Journal of Computer Assisted Learning*, *37*(1), 183–196. https://doi.org/10.1111/jcal.12478

Anker-Hansen, J., & Andreé, M. (2019). In pursuit of authenticity in science education. *Nordic Studies in Science Education*, *15*(1), 54–66. https://doi.org/10.5617/nordina.4723

Apoki, U. C., Ali Hussein, A. M., Al-Chalabi, H. K. M., Badica, C., & Mocanu, M. L. (2022). The Role of Pedagogical Agents in Personalised Adaptive Learning: A Review. *Sustainability (Switzerland)*, *14*(11), 1–17. https://doi.org/10.3390/su14116442

Araiza-Alba, P., Keane, T., Matthews, B., Simpson, K., Strugnell, G., Chen, W. S., & Kaufman, J. (2021). The potential of 360-degree virtual reality videos to teach water-safety skills to children. *Computers & Education*, *163*(November 2020), 104096. https://doi.org/10.1016/j.compedu.2020.104096

Ardiny, H., & Khanmirza, E. (2018). The Role of AR and VR Technologies in Education Developments: Opportunities and Challenges. *2018 6th RSI International Conference on Robotics and Mechatronics (IcRoM)*, 482–487. https://doi.org/10.1109/ICRoM.2018.8657615

Arnaldi, B., Guitton, P., & Moreau, G. (2018). *Virtual Reality and Augmented Reality: Myths and Realities* (1st ed.). Wiley-IEEE Press.

Bandura, A. (1977). Social learning theory. In *Social learning theory.* Prentice-Hall.

Bayliss, A. P., Schuch, S., & Tipper, S. P. (2010). Gaze cueing elicited by emotional faces is influenced by affective context. *Visual Cognition*, *18*(8), 1214–1232. https://doi.org/10.1080/13506285.2010.484657

Beege, M., Nebel, S., Schneider, S., & Rey, G. D. (2019). Social entities in educational videos: Combining the effects of addressing and professionalism. *Computers in Human Behavior*, *93*, 40–52. https://doi.org/10.1016/j.chb.2018.11.051

Beege, M., Nebel, S., Schneider, S., & Rey, G. D. (2023). The effect of microlevel and macrolevel signaling on learning with 360° videos. *Applied Cognitive Psychology*, *37*(1), 232–246. https://doi.org/10.1002/acp.4023

Bender, A. D., Filmer, H. L., Naughtin, C. K., & Dux, P. E. (2017). Dynamic, continuous multitasking training leads to task-specific improvements but does not transfer across action selection tasks. *Npj Science of Learning*, *2*(1). https://doi.org/10.1038/s41539-017-0015-4

Bender, S. (2019). Headset attentional synchrony: tracking the gaze of viewers watching narrative virtual reality. *Media Practice and Education*, *20*(3), 277–296. https://doi.org/10.1080/25741136.2018.1464743

Betz, A. (2018). The influence of the learning environment on learners' (perceived) authenticity of science communication and on their situational interest. *Unterrichtswissenschaft*, *46*(3), 261–278. https://doi.org/10.1007/s42010-018-0021-0

BGHM. (2019). *DGUV Regel 109-606 Branche Tischler- und Schreinerhandwerk* (Issue April). Berufsgenossenschaft Holz und Metall. https://www.bghm.de/fileadmin/user_upload/Arbeitsschuetzer/Gesetze_Vorschriften/Regeln/109-606.pdf

Bian, Y. (2022). Motivation Effect of Animated Pedagogical Agent's Personality and Feedback Strategy Types on Learning in Virtual Training Environment. *Virtual Reality and Intelligent Hardware*, *4*(2), 153–172. https://doi.org/10.1016/j.vrih.2021.11.001

Bloom, B. S. (1956). *Taxonomy of Educational Objectives: Handbook II*. David McKay.

Blumer, H. (1969). *Symbolic interactionism; perspective and method*. Prentice-Hall.

Blurton, S. P., Greenlee, M. W., & Gondan, M. (2015). Cross-modal cueing in audiovisual spatial attention. *Attention, Perception, and Psychophysics*, *77*(7), 2356–2376. https://doi.org/10.3758/s13414-015-0920-7

Botch, T. L., Garcia, B. D., Choi, Y. B., Feffer, N., & Robertson, C. E. (2023). Active visual search in naturalistic environments reflects individual differences in classic visual search performance. *Scientific Reports*, *13*(1), 1–9. https://doi.org/10.1038/s41598-023-27896-7

Boucheix, J. M., Lowe, R. K., Putri, D. K., & Groff, J. (2013). Cueing animations: Dynamic signaling aids information extraction and comprehension. *Learning and Instruction*, *25*, 71–84. https://doi.org/10.1016/j.learninstruc.2012.11.005

Bratzke, D., Ulrich, R., Rolke, B., Schroter, H., Jentzsch, I., & Leuthold, H. (2008). Motor limitation in dual-task processing with different effectors. *Quarterly Journal of Experimental Psychology*, *61*(9), 1385–1399. https://doi.org/10.1080/17470210701536856

Brucker, B., Pardi, G., Uehlin, F., Moosmann, L., Lachmair, M., Halfmann, M., & Gerjets, P. (2024). How Learners' Visuospatial Ability and Different Ways of Changing the Perspective Influence Learning About Movements in Desktop and Immersive Virtual Reality Environments. In *Educational Psychology Review* (Vol. 36, Issue 3). Springer US. https://doi.org/10.1007/s10648-024-09895-w

Burgoon, J. K. (1993). Interpersonal Expectations, Expectancy Violations, and Emotional Communication. *Journal of Language and Social Psychology*, *12*(1–2), 30–48. https://doi.org/10.1177/0261927X93121003

Buttussi, F., & Chittaro, L. (2018). Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. *IEEE Transactions on Visualization and Computer Graphics*, *24*(2), 1063–1076. https://doi.org/10.1109/TVCG.2017.2653117

Castro-Alonso, J. C., Wong, R. M., Adesope, O. O., & Paas, F. (2021). Effectiveness of Multimedia Pedagogical Agents Predicted by Diverse Theories: a Meta-Analysis. *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09587-1

Catrysse, L., Kerckhoffs, A., & Jarodzka, H. (2023). Behind the scenes: Impact of virtual backgrounds in educational videos on visual processing and learning outcomes. *Journal of Eye Movement Research*, *16*(3). https://doi.org/10.16910/jemr.16.3.4

Chacón-Candia, J. A., Román-Caballero, R., Aranda-Martín, B., Casagrande, M., Lupiáñez, J., & Marotta, A. (2023). Are there quantitative differences between eye-gaze and arrow cues? A meta-analytic answer to the debate and a call for qualitative differences. *Neuroscience & Biobehavioral Reviews*, *144*, 104993. https://doi.org/https://doi.org/10.1016/j.neubiorev.2022.104993

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual Reality Sickness: A Review of Causes and Measurements. *International Journal of Human-Computer Interaction*, *36*(17), 1658–1682. https://doi.org/10.1080/10447318.2020.1778351

Cheng, K.-H., & Tsai, C.-C. (2019). A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers & Education*, *140*, 103600. https://doi.org/10.1016/j.compedu.2019.103600

Clinch, S., Cortis Mack, C., Ward, G., & Steeds, M. (2021). *Technology-Mediated Memory Impairment BT - Technology-Augmented Perception and Cognition* (T. Dingler & E. Niforatos (eds.); pp. 71–124). Springer International Publishing. https://doi.org/10.1007/978-3-030-30457-7_4

Conrad, M., Kablitz, D., & Schumann, S. (2024). Learning effectiveness of immersive virtual reality in education and training: A systematic review of findings. *Computers & Education: X Reality*, *4*(September 2023), 100053. https://doi.org/10.1016/j.cexr.2024.100053

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, *94*(2), 428–434. https://doi.org/10.1037/0022-0663.94.2.428

Cummings, J. J., & Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, *19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Dahmani, L., & Bohbot, V. D. (2020). Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-62877-0

Dai, L., Jung, M. M., Postma, M., & Louwerse, M. M. (2022). A systematic review of pedagogical agent research: Similarities, differences and unexplored aspects. *Computers and Education*, *190*(April), 104607. https://doi.org/10.1016/j.compedu.2022.104607

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, *41*(1), 10–32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Damhorst, M. L. (1990). In Search of a Common Thread: Classification of Information Communicated Through Dress. *Clothing and Textiles Research Journal*, *8*(2), 1–12. https://doi.org/10.1177/0887302X9000800201

Davis, R. O. (2018). The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis. *Educational Research Review*, *24*, 193–209. https://doi.org/10.1016/j.edurev.2018.05.002

Davis, R. O., Park, T., & Vincent, J. (2023). A Meta-Analytic Review on Embodied Pedagogical Agent Design and Testing Formats. *Journal of Educational Computing Research*, *61*(1), 30–67. https://doi.org/10.1177/07356331221100556

de Bruin, A. B. H., Biwer, F., Hui, L., Onan, E., David, L., & Wiradhany, W. (2023). Worth the Effort: the Start and Stick to Desirable Difficulties (S2D2) Framework. In *Educational Psychology Review* (Vol. 35, Issue 2). Springer. https://doi.org/10.1007/s10648-023-09766-w

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology*, *21*(6), 731–746. https://doi.org/10.1002/acp.1346

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, *20*(2), 111–122. https://doi.org/10.1016/j.learninstruc.2009.02.010

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2011). Attention cueing in an instructional animation: The role of presentation speed. *Computers in Human Behavior*, *27*(1), 41–45. https://doi.org/10.1016/j.chb.2010.05.010

Decker, D., & Merkt, M. (2023). Do a pedagogical agent's clothing and an animated video's setting affect learning? *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1205338

Decker, D., & Merkt, M. (2024). Shake it or light it! The effects of cueing in desktop-VR learning environments on search time and learning. *Journal of Computer Assisted Learning*, *40*(3), 1201–1217. https://doi.org/https://doi.org/10.1111/jcal.12945

Dolsak, W., & Reich, C. (2022). Virtual reality simulator for pipe umbrella installation works. *Geomechanik Und Tunnelbau*, *15*(2), 233–238. https://doi.org/10.1002/geot.202100062

Dunbar, N. E., & Segrin, C. (2012). Clothing and Teacher Credibility: An Application of Expectancy Violations Theory. *ISRN Education*, *2012*, 1–12.

https://doi.org/10.5402/2012/140517

Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., & Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron*, *63*(1), 127–138. https://doi.org/10.1016/j.neuron.2009.06.005

Fombona-Pascual, A., Fombona, J., & Vázquez-Cano, E. (2022). VR in chemistry, a review of scientific research on advanced atomic/molecular visualization. *Chemistry Education Research and Practice*, *23*(2), 300–312. https://doi.org/10.1039/d1rp00317h

Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., & Andersen, S. A. W. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surgical Endoscopy*, *34*(3), 1244–1252. https://doi.org/10.1007/s00464-019-06887-8

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Fromm, C. A. (2023). Learning to see and hear in 3D: Virtual reality as a platform for multisensory perceptual learning [ProQuest Information & Learning]. In *Dissertation Abstracts International: Section B: The Sciences and Engineering* (Vol. 84, Issues 6-B). http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx%3Fdirect%3Dtrue%26db%3Dpsyh%26AN%3D2023-38658-150%26site%3Dehost-live

Gardony, A. L., Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2013). How Navigational Aids Impair Spatial Memory: Evidence for Divided Attention. *Spatial Cognition and Computation*, *13*(4), 319–350. https://doi.org/10.1080/13875868.2013.792821

Gardony, A. L., Brunyé, T. T., & Taylor, H. A. (2015). Navigational Aids and Spatial

Memory Impairment: The Role of Divided Attention. *Spatial Cognition and Computation*, *15*(4), 246–284. https://doi.org/10.1080/13875868.2015.1059432

Gazit, N., Ben-Gal, G., & Eliashar, R. (2024). Development and validation of an objective virtual reality tool for assessing technical aptitude among potential candidates for surgical training. *BMC Medical Education*, *24*(1), 1–17. https://doi.org/10.1186/s12909-024-05228-1

Grant, S. C., & Magee, L. E. (1998). Contributions of proprioception to navigation in virtual environments. *Human Factors*, *40*(3), 489–497. https://doi.org/10.1518/001872098779591296

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2023). Mutual interplay between cognitive offloading and secondary task performance. *Psychonomic Bulletin and Review*, *0123456789*. https://doi.org/10.3758/s13423-023-02312-3

Grivokostopoulou, F., Kovas, K., & Perikos, I. (2020). The effectiveness of embodied pedagogical agents and their impact on students learning in virtual worlds. *Applied Sciences (Switzerland)*, *10*(5), 1–14. https://doi.org/10.3390/app10051739

Grogorick, S., Albuquerque, G., & Magnor, M. (2018). Comparing Unobtrusive Gaze Guiding Stimuli in Head-Mounted Displays. *Proceedings - International Conference on Image Processing, ICIP*, 2805–2809. https://doi.org/10.1109/ICIP.2018.8451784

Gurney, D. J., Howlett, N., Pine, K., Tracey, M., & Moggridge, R. (2017). Dressing up posture: The interactive effects of posture and clothing on competency judgements. *British Journal of Psychology*, *108*(2), 436–451. https://doi.org/10.1111/bjop.12209

Gurung, R. A. R., Kempen, L., Klemm, K., Senn, R., & Wysocki, R. (2014). Dressed to

Present: Ratings of Classroom Presentations Vary With Attire. *Teaching of Psychology*, *41*(4), 349–353. https://doi.org/10.1177/0098628314549710

Hartmans, C. (2013). The Doctor's New Clothes: Professional or Fashionable? *Primary Health Care: Open Access*, *03*(03). https://doi.org/10.4172/2167-1079.1000145

Hazeltine, E., Teague, D., & Ivry, R. B. (2002). Simultaneous dual-task performance reveals parallel response selection after practice. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 527–545. https://doi.org/10.1037/0096-1523.28.3.527

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, *6*(1), 27–54. https://doi.org/10.1016/j.edurev.2010.07.004

Hejtmánek, L., Oravcová, I., Motýl, J., Horáček, J., & Fajnerová, I. (2018). Spatial knowledge impairment after GPS guided navigation: Eye-tracking study in a virtual town. *International Journal of Human Computer Studies*, *116*(May 2017), 15–24. https://doi.org/10.1016/j.ijhcs.2018.04.006

Hooge, I., & Camps, G. (2013). Scan path entropy and arrow plots: Capturing scanning behavior of multiple observers. *Frontiers in Psychology*, *4*(DEC), 1–10. https://doi.org/10.3389/fpsyg.2013.00996

Hopkins, K., Kass, S. J., Blalock, L. D., & Brill, J. C. (2017). Effectiveness of auditory and tactile crossmodal cues in a dual-task visual and auditory scenario. *Ergonomics*, *60*(5), 692–700. https://doi.org/10.1080/00140139.2016.1198495

Hornof, A. J., Zhang, Y., & Halverson, T. (2010). Knowing where and when to look in a time-critical multimodal dual task. *Conference on Human Factors in Computing Systems - Proceedings*, *3*, 2103–2112. https://doi.org/10.1145/1753326.1753647

Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of*

Environmental Psychology, *28*(1), 74–82. https://doi.org/10.1016/j.jenvp.2007.09.002

Jackson, K. M., Shaw, T. H., & Helton, W. S. (2023a). Evaluating the dual-task decrement within a simulated environment: Word recall and visual search. *Applied Ergonomics*, *106*(August 2022), 103861. https://doi.org/10.1016/j.apergo.2022.103861

Jackson, K. M., Shaw, T. H., & Helton, W. S. (2023b). The effects of dual-task interference on visual search and verbal memory. *Ergonomics*, *66*(1), 125–135. https://doi.org/10.1080/00140139.2022.2061053

Jiang, L., Xu, M., Liu, T., Qiao, M., & Wang, Z. (2018). DeepVS: A Deep Learning Based Video Saliency Prediction Approach. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (pp. 625–642). Springer International Publishing. https://doi.org/10.1007/978-3-030-01264-9_37

Johnson-Glenberg, M. C. (2019). *The Necessary Nine: Design Principles for Embodied VR and Active Stem Education*. Springer Singapore. https://doi.org/10.1007/978-981-13-8265-9_5

Johnson-Glenberg, M. C., Bartolomea, H., & Kalina, E. (2021). Platform is not destiny: Embodied learning effects comparing 2D desktop to 3D virtual reality STEM experiences. *Journal of Computer Assisted Learning*, *37*(5), 1263–1284. https://doi.org/10.1111/jcal.12567

Kang, Y. G., Song, H. D., Yun, H., & Jo, Y. (2022). The effect of virtual reality media characteristics on flow and learning transfer in job training: The moderating effect of presence. *Journal of Computer Assisted Learning*, *38*(6), 1674–1685. https://doi.org/10.1111/jcal.12702

Khokhar, A., & Borst, C. (2022). Modifying Pedagogical Agent Spatial Guidance Sequences to Respond to Eye-Tracked Student Gaze in VR. *Proceedings - SUI 2022: ACM Conference on Spatial User Interaction*. https://doi.org/10.1145/3565970.3567697

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). *Showing face in video*

instruction. 2095–2102. https://doi.org/10.1145/2556288.2557207

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*(NOV), 1–18. https://doi.org/10.3389/fpsyg.2017.01997

Kodžoman, D. (2019). The psychology of clothing. *Textile & Leather Review*, *2*(2), 90–103. https://doi.org/10.31881/tlr.2019.22

Landau, M., Comeaux, M., Mortell, T., Boyle, R., Imbrescia, K., & Chaffin, A. E. (2024). Characterizing the Untapped Potential of Virtual Reality in Plastic and Reconstructive Surgical Training: A Systematic Review on Skill Transferability. *JPRAS Open*, *41*, 295–310. https://doi.org/10.1016/j.jpra.2024.06.015

Lange, D., Stratmann, T. C., Gruenefeld, U., & Boll, S. (2020). HiveFive: Immersion Preserving Attention Guidance in Virtual Reality. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, *April*, 1–13. https://doi.org/10.1145/3313831.3376803

Leder, J., Horlitz, T., Puschmann, P., Wittstock, V., & Schütz, A. (2019). Comparing immersive virtual reality and powerpoint as methods for delivering safety training: Impacts on risk perception, learning, and decision making. *Safety Science*, *111*(February 2018), 271–286. https://doi.org/10.1016/j.ssci.2018.07.021

Lee, G., Lee, D. Y., Su, G.-M., & Manocha, D. (2024). "May I Speak?": Multi-Modal Attention Guidance in Social VR Group Conversations. *IEEE Transactions on Visualization and Computer Graphics*, *30*(5), 2287–2297. https://doi.org/10.1109/TVCG.2024.3372119

Lee, K. M. (2004). Presence, Explicated. *Communication Theory*, *14*(1), 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lester, J. C., Barlow, S. T., Converse, S. A., Stone, B. A., Kahler, S. E., & Bhogal, R. S. (1997). Persona effect: Affective impact of animated pedagogical agents. *Conference on*

Human Factors in Computing Systems - Proceedings, 359–366.

Li, Wenhao, Wang, C., Wang, W., Feng, Q., & Liu, B. (2024). Learning in the spherical video-based virtual reality context: effects of immersion level and spatial ability on learning performance and experience. *Interactive Learning Environments*, 1–22. https://doi.org/10.1080/10494820.2024.2367019

Li, Wenjing, Wang, F., & Mayer, R. E. (2023). How to guide learners' processing of multimedia lessons with pedagogical agents. *Learning and Instruction*, *84*, 101729. https://doi.org/https://doi.org/10.1016/j.learninstruc.2022.101729

Li, Wenjing, Wang, F., & Mayer, R. E. (2024). Increasing the realism of on-screen embodied instructors creates more looking but less learning. *British Journal of Educational Psychology*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/bjep.12677

Li, Wenjing, Wang, F., Mayer, R. E., & Liu, H. (2019). Getting the point: Which kinds of gestures by pedagogical agents improve multimedia learning? *Journal of Educational Psychology*, *111*(8), 1382–1395. https://doi.org/10.1037/edu0000352

Li, Wenjing, Wang, F., Mayer, R., & Liu, T. (2022). Animated pedagogical agents enhance learning outcomes and brain activity during learning. *Journal of Computer Assisted Learning*, *38*. https://doi.org/10.1111/jcal.12634

Lin, L., & Atkinson, R. K. (2011). Using animations and visual cueing to support learning of scientific concepts and processes. *Computers and Education*, *56*(3), 650–658. https://doi.org/10.1016/j.compedu.2010.10.007

Lin, Y.-C., Chang, Y.-J., Hu, H.-N., Cheng, H.-T., Huang, C.-W., & Sun, M. (2017). Tell Me Where to Look. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2535–2545. https://doi.org/10.1145/3025453.3025757

Liu, R., Xu, X., Yang, H., Li, Z., & Huang, G. (2022). Impacts of Cues on Learning and Attention in Immersive 360-Degree Video: An Eye-Tracking Study. *Frontiers in Psychology*, *12*, 1–11. https://doi.org/10.3389/fpsyg.2021.792069

Lukavsky, J., Butler, S., & Harden, A. J. (1995). Perceptions of an Instructor: Dress and Students' Characteristics. *Perceptual and Motor Skills*, *81*(1), 231–240. https://doi.org/10.2466/pms.1995.81.1.231

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, *113*(4), 719–735. https://doi.org/10.1037/edu0000473

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, *66*(5), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, *72*, 276–285. https://doi.org/10.1016/j.chb.2017.02.066

Makransky, G., Mayer, R., Nøremølle, A., Cordoba, A. L., Wandall, J., & Bonde, M. (2020). Investigating the feasibility of using assessment and explanatory feedback in desktop virtual reality simulations. *Educational Technology Research and Development*, *68*(1), 293–317. https://doi.org/10.1007/s11423-019-09690-3

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09586-2

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, *60*(1), 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Marsh, T., & Smith, S. P. (2001). Guiding User Navigation in Virtual Environments Using Awareness of Virtual Off-Screen Space. *Proceedings of the Workshop on Guiding Users through Interactive Experiences—Usability Centred Design and Evaluation of Virtual 3D Environments*, 149–154. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.4719

Martha, A. S. D., & Santoso, H. B. (2019). The design and impact of the pedagogical agent: A systematic literature review. *Journal of Educators Online*, *16*(1). https://doi.org/10.9743/jeo.2019.16.1.8

Mayer, R. E. (2014). The Cambridge Handbook of Multimedia Learning. In R. E. Mayer (Ed.), *Cambridge Handbooks in Psychology* (2nd ed.). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781139547369

Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia Learning in an Interactive Self-Explaining Environment: What Works in the Design of Agent-Based Microworlds? *Journal of Educational Psychology*, *95*(4), 806–812. https://doi.org/10.1037/0022-0663.95.4.806

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology*, *95*(2), 419–425. https://doi.org/10.1037/0022-0663.95.2.419

McKay, K. T., Grainger, S. A., Coundouris, S. P., Skorich, D. P., Phillips, L. H., & Henry, J. D. (2021). Visual attentional orienting by eye gaze: A meta-analytic review of the gaze-cueing effect. *Psychological Bulletin*, *147*(12), 1269–1289. https://doi.org/10.1037/bul0000353

Merkt, M., Lux, S., Hoogerheide, V., van Gog, T., & Schwan, S. (2020). A change of scenery: Does the setting of an instructional video affect learning? *Journal of Educational Psychology*, *112*(6), 1223–1283. https://doi.org/10.1037/edu0000414

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive

processes and multiple-task performance: Part I. Basic mechanisms. *Psychological Review*, *104*(1), 3–65. https://doi.org/10.1037/0033-295X.104.1.3

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, *140*, 103603. https://doi.org/https://doi.org/10.1016/j.compedu.2019.103603

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999-2009). *Computers and Education*, *56*(3), 769–780. https://doi.org/10.1016/j.compedu.2010.10.020

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition*, *21*(1), 59–68. https://doi.org/10.1016/j.concog.2011.08.010

Moreno, R. (2012). Multimedia Learning with Animated Pedagogical Agents. *The Cambridge Handbook of Multimedia Learning*, 507–524. https://doi.org/10.1017/cbo9780511816819.032

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Moritz, J., Meyerhoff, H. S., & Schwan, S. (2020). Control over spatial representation format enhances information extraction but prevents long-term learning. *Journal of Educational Psychology*, *112*(1), 148–165. https://doi.org/10.1037/edu0000364

Obourdin, G., de Maeyer, S., & Van den Bossche, P. (2024). Unlocking the power of immersive learning: The FAIRI instructional design proposition for adaptive immersive virtual reality. *Computers and Education: X Reality*, *5*(September), 100084. https://doi.org/10.1016/j.cexr.2024.100084

Olmos-Raya, E., Ferreira-Cavalcanti, J., Contero, M., Castellanos, M. C., Giglioli, I. A. C., & Alcañiz, M. (2018). Mobile virtual reality as an educational platform: A pilot study on

the impact of immersion and positive emotion induction in the learning process. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(6), 2045–2057. https://doi.org/10.29333/ejmste/85874

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, *110*(6), 785–797. https://doi.org/10.1037/edu0000241

Parong, J., & Mayer, R. E. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*, *37*(1), 226–241. https://doi.org/10.1111/jcal.12482

Pashler, H. (1994). Dual-Task Interference in Simple Tasks: Data and Theory. *Psychological Bulletin*, *116*(2), 220–244. https://doi.org/10.1037//0033-2909.116.2.220

Pejsa, T., Andrist, S., Gleicher, M., & Mutlu, B. (2015). Gaze and attention management for embodied conversational agents. *ACM Transactions on Interactive Intelligent Systems*, *5*(1), 1–34. https://doi.org/10.1145/2724731

Peng, N., Song, L., Wu, L., Wang, H., & Xue, C. (2022). Cueing effects of colour on attention management in multiple-view visualisations: evidence from eye-tracking by using a dual-task paradigm. *Behaviour and Information Technology*, *41*(8), 1652–1670. https://doi.org/10.1080/0144929X.2021.1895318

Petersen, G. B. g., Mottelson, A., & Makransky, G. (2021). Pedagogical agents in educational vr: An in the wild study. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3411764.3445760

Porte, L., Boucheix, J. M., Rapet, L., Drai-Zerbib, V., & Martinez, J. L. (2024). Immersive virtual reality for learning about ecosystems: effect of two signaling levels and feedback on action decisions. *Frontiers in Psychology*, *15*(April). https://doi.org/10.3389/fpsyg.2024.1359071

Queiroz, A. C. M., Fauville, G., Herrera, F., Leme, M. I. da S., & Bailenson, J. N. (2022). Do Students Learn Better With Immersive Virtual Reality Videos Than Conventional

Videos? A Comparison of Media Effects With Middle School Girls. *Technology, Mind, and Behavior*, *3*(3). https://doi.org/10.1037/tmb0000082

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778. https://doi.org/https://doi.org/10.1016/j.compedu.2019.103778

Renganayagalu, S. kumar, Mallam, S. C., & Nazir, S. (2021). Effectiveness of VR Head Mounted Displays in Professional Training: A Systematic Review. In *Technology, Knowledge and Learning* (Vol. 26, Issue 4). Springer Netherlands. https://doi.org/10.1007/s10758-020-09489-9

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, *7*(3), 216–237. https://doi.org/10.1016/j.edurev.2012.05.003

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, *17*, 19–36. https://doi.org/10.1016/j.edurev.2015.12.003

Rothe, S., & Hußmann, H. (2018). Guiding the Viewer in Cinematic Virtual Reality by Diegetic Cues. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 10850 LNCS* (pp. 101–117). https://doi.org/10.1007/978-3-319-95270-3_7

Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educational Research Review*, *23*, 1–24. https://doi.org/10.1016/j.edurev.2017.11.001

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The Cognitive-Affective-Social Theory of Learning in digital Environments (CASTLE). *Educational Psychology Review*, *34*(1), 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schroeder, N., & Adesope, O. (2012). A case for the use of pedagogical agents in online learning environments. *Journal of Teaching and Learning with Technology*, *1*(2), 43–47.

Schüttler, T., Watzka, B., Girwidz, R., & Ertl, B. (2021). Die Wirkung der Authentizität von Lernort und Laborgeräten auf das situationale Interesse und die Relevanzwahrnehmung beim Besuch eines naturwissenschaftlichen Schülerlabors. *Zeitschrift Für Didaktik Der Naturwissenschaften*. https://doi.org/10.1007/s40573-021-00128-z

Seibert, J., & Shafer, D. M. (2018). Control mapping in virtual reality: effects on spatial presence and controller naturalness. *Virtual Reality*, *22*(1), 79–88. https://doi.org/10.1007/s10055-017-0316-1

Shi, L., Huang, F. C., Lopes, W., Matusik, W., & Luebke, D. (2017). Near-eye light field holographic rendering with spherical waves for wide field of view interactive 3D computer graphics. *ACM Transactions on Graphics*, *36*(6). https://doi.org/10.1145/3130800.3130832

Shu, Y., Huang, Y. Z., Chang, S. H., & Chen, M. Y. (2019). Do virtual reality head-mounted displays make a difference? A comparison of presence and self-efficacy between head-mounted displays and desktop computer-facilitated virtual environments. *Virtual Reality*, *23*(4), 437–446. https://doi.org/10.1007/s10055-018-0376-x

Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers and Education*, *188*(May). https://doi.org/10.1016/j.compedu.2022.104564

Skulmowski, A. (2023). The Cognitive Architecture of Digital Externalization. *Educational Psychology Review*, *35*(4), 1–21. https://doi.org/10.1007/s10648-023-09818-1

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers Robotics AI*, *3*(DEC), 1–47. https://doi.org/10.3389/frobt.2016.00074

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators*

and Virtual Environments, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Smith, S. P., & Marsh, T. (2004). Evaluating Design Guidelines for Reducing User Disorientation in a Desktop Virtual Environment. *Virtual Real.*, *8*(1), 55–62.

Sondermann, C., Huff, M., & Merkt, M. (2024). Distracted by a talking head? An eye tracking study on the effects of instructor presence in learning videos with animated graphic slides. *Learning and Instruction*, *91*(January), 101878. https://doi.org/10.1016/j.learninstruc.2024.101878

Souchet, A. D., Philippe, S., Lourdeaux, D., & Leroy, L. (2022). Measuring Visual Fatigue and Cognitive Load via Eye Tracking while Learning with Virtual Reality Head-Mounted Displays: A Review. *International Journal of Human-Computer Interaction*, *38*(9), 801–824. https://doi.org/10.1080/10447318.2021.1976509

Speicher, M., Rosenberg, C., Degraen, D., Daiber, F., & Krúger, A. (2019). Exploring Visual Guidance in 360-degree Videos. *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, 1–12. https://doi.org/10.1145/3317697.3323350

Stefanucci, J. K., Brickler, D., Finney, H. C., Wilson, E., Drew, T., & Creem-Regehr, S. H. (2022). Effects of simulated augmented reality cueing in a virtual navigation task. *Frontiers in Virtual Reality*, *3*(September), 1–16. https://doi.org/10.3389/frvir.2022.971310

Strobach, T., Frensch, P. A., Soutschek, A., & Schubert, T. (2012). Investigation on the improvement and transfer of dual-task coordination skills. *Psychological Research*, *76*(6), 794–811. https://doi.org/10.1007/s00426-011-0381-0

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Intrinsic and Extraneous Cognitive Load. In *Cognitive Load Theory* (pp. 57–69). Springer New York. https://doi.org/10.1007/978-1-4419-8126-4_5

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and

Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Tusher, H. M., Mallam, S., & Nazir, S. (2024). A Systematic Review of Virtual Reality Features for Skill Training. In *Technology, Knowledge and Learning* (Vol. 29, Issue 2). Springer Netherlands. https://doi.org/10.1007/s10758-023-09713-2

van der Haak, B. (2014). Digital Amnesia at Work. *Kaspersky Lab*. https://media.kasperskydaily.com/wp-content/uploads/sites/89/2016/03/05232337/Digital-Amnesia-at-work-the-risks-and-rewards-of-forgetting-in-business-report.pdf

van Gog, T. (2021). The Signaling (or Cueing) Principle in Multimedia Learning. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 221–230). Cambridge University Press. https://doi.org/DOI: 10.1017/9781108894333.022

Veletsianos, G. (2010). Contextually relevant pedagogical agents: Visual appearance, stereotypes, and first impressions and their impact on learning. *Computers and Education*, *55*(2), 576–585. https://doi.org/10.1016/j.compedu.2010.02.019

Volkmann, T., Wessel, D., Jochems, N., & Franke, T. (2018). German Translation of the Multimodal Presence Scale. *Dachselt, R. & Weber, G. (Hrsg.), Mensch Und Computer 2018 - Tagungsband.*, *September 2018*. http://dx.doi.org/https://doi.org/10.18420/muc2018-mci-0428

Wang, W.-S., Pedaste, M., Lin, C.-J., Lee, H.-Y., Huang, Y.-M., & Wu, T.-T. (2024). Signaling feedback mechanisms to promoting self-regulated learning and motivation in virtual reality transferred to real-world hands-on tasks. *Interactive Learning Environments*, 1–16. https://doi.org/10.1080/10494820.2024.2331151

Wang, Y., Gong, S., Cao, Y., & Fan, W. (2023). The power of affective pedagogical agent and self-explanation in computer-based learning. *Computers & Education*, *195*, 104723.

https://doi.org/https://doi.org/10.1016/j.compedu.2022.104723

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Wu, Y., & Wang, Q. (2024). Dual-task interference: Bottleneck constraint or capacity sharing? Evidence from automatic and controlled processes. *Attention, Perception, and Psychophysics*, *86*(3), 815–827. https://doi.org/10.3758/s13414-024-02854-1

Zhang, H., Zhang, Y., Xu, T., & Zhou, Y. (2023). Effects of VR instructional approaches and textual cues on performance, cognitive load, and learning experience. *Educational Technology Research and Development*, *0123456789*. https://doi.org/10.1007/s11423-023-10313-1

Zhu, Y., Zhai, G., Min, X., & Zhou, J. (2020). The Prediction of Saliency Map for Head and Eye Movements in 360 Degree Images. *IEEE Transactions on Multimedia*, *22*(9), 2331–2344. https://doi.org/10.1109/TMM.2019.2957986

Appendix A

Manuscript 1

Decker, D., & Merkt, M. (2023). Do a pedagogical agent's clothing and an animated video's setting affect learning? *Frontiers in Psychology*, 14. https://doi.org/10.3389/fpsyg.2023.1205338

Author	Author Position	Scientific Ideas %	Data generation %	Analysis and interpretation %	Paper writing %
Daniela Decker	1	70	100	80	80
Martin Merkt	2	30	0	20	20

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 26 September 2023
DOI 10.3389/fpsyg.2023.1205338

OPEN ACCESS

EDITED BY
Laura M. Morett,
University of Missouri, United States

REVIEWED BY
Maik Beege,
University of Education Freiburg, Germany
Xinghua Wang,
Qingdao University, China

*CORRESPONDENCE
Daniela Decker
✉ decker@die-bonn.de

RECEIVED 13 April 2023
ACCEPTED 12 September 2023
PUBLISHED 26 September 2023

CITATION
Decker D and Merkt M (2023) Do a pedagogical agent's clothing and an animated video's setting affect learning? *Front. Psychol.* 14:1205338. doi: 10.3389/fpsyg.2023.1205338

Do a pedagogical agent's clothing and an animated video's setting affect learning?

Daniela Decker* and Martin Merkt

German Institute for Adult Education – Leibniz Centre for Lifelong Learning, Bonn, Germany

Pedagogical agents are often used to enhance social cues in learning materials. The inclusion of pedagogical agents raises several design questions, for example on what kind of clothing the agent should wear. Further, it is not yet clear how the setting of an animated learning video (i.e., the digitally created background) affects learning. In an online experiment (N = 200), we investigated whether creating thematically appropriate clothing and setting has some added value in that it improves learning outcomes in comparison to more neutral assets. Whereas all participants acquired knowledge from the animated video, there were no main effects of clothing and setting for any of the dependent variables, but an interaction for learning outcomes ($\eta_p^2 = 0.02$), indicating that the appropriately dressed agent worked better combined with the inappropriate setting than with the appropriate setting. Overall, given those non-significant main effects and the small effect size of the interaction, there seem to be some degrees of freedom for designers of pedagogical agents in animated learning videos. However, these degrees of freedom may be limited to at least moderate (i.e., neutral) levels of appropriateness.

KEYWORDS

pedagogical agent, adult learning, distance education, online learning, clothing, setting

1. Introduction

Learning with (animated) online videos can take place in a variety of contexts, for example as additional learning material in formal education or in informal learning contexts, where learners choose learning content self-directed and decide for themselves how they want to learn with these materials (Lange, 2018). Given this increasing popularity among learners and the growing opportunities for educators to create these videos, it is important to understand how the decisions to design these materials affect learning. Regarding the increasing possibilities to include humanlike virtual characters into animated learning materials (i.e., pedagogical agents; Craig et al., 2002; Apoki et al., 2022), the design of such agents becomes another factor to consider in terms of its impact on learning. Whereas it is evident that the inclusion of a virtual pedagogical agent may serve as a social cue (Lester et al., 1997; Li et al., 2022), there are still some open questions regarding the specific design of the pedagogical agent as well as the content-appropriateness of the setting (i.e., the digitally created background) in that the pedagogical agent is integrated. In this manuscript, we will address whether designing thematically appropriate clothing and an appropriate setting for a pedagogical agent in an animated learning video benefits learners' evaluations of the learning materials and learning outcomes. Before we go into detail about the empirical background

regarding the pedagogical agent's clothing and the setting of the learning materials, we will give a brief overview of potential effects of pedagogical agents in general.

1.1. The importance of social cues for learning

Influential theories such as the cognitive load theory (CLT) (Sweller et al., 1998; Sweller, 2011) and the cognitive theory of multimedia learning (CTML) (Mayer, 2014) still have a strong focus on the cognitive processes that underly successful learning. Most importantly, both theories share the assumption that learners' cognitive resources are limited, and learning materials should be designed so that they do not result in cognitive overload. The CTML further states that learners must actively select, organize, and integrate incoming information to develop comprehensive mental representations in long-term memory. In such a purely cognitive approach, a pedagogical agent may be considered an interesting, but irrelevant addition to the learning materials, thus constituting a seductive detail that may increase extraneous cognitive load and thus hinder learning (for an overview, see Rey, 2012). However, extending these primarily cognitive theories, theories such as the social agency theory (Mayer et al., 2003; Mayer, 2014) emerged that emphasize the role of social cues in the learning process (also see social learning theory; Bandura, 1977). Further, the cognitive-affective-social theory of learning in digital environments (CASTLE; Schneider et al., 2022), which integrates a variety of theories on social cues, distinguishes between different types of social cues which may activate social schemas in learners resulting in increased (para-)social and metacognitive processes. These processes should subsequently affect the information selection, as well as the further processing in the working memory, which might also influence the building of mental models in long-term memory. In this regard, CASTLE (Schneider et al., 2022) emphasizes the use pedagogical agents as a means to provide both verbal and visual social cues. Therefore, including a pedagogical agent to provide social cues should have beneficial effects on learning due to an increased social response (Lester et al., 1997), which is reflected in increased brain activity during learning in areas of the brain that are usually associated with social processes (Li et al., 2022). Moreover, social cues provided by pedagogical agents might influence learning indirectly through factors such as motivation (Heidig and Clarebout, 2011; Bian, 2022; Sikström et al., 2022; Wang et al., 2023), perceived learning experience (Grivokostopoulou et al., 2020), and the perceived expertise of a pedagogical agent (Nguyen, 2023).

1.2. Use of pedagogical agents

Regarding the influence of pedagogical agents on recall, transfer, and motivation, Heidig and Clarebout (2011) suggested the pedagogical agents conditions of use (PACU) framework, due to the rather heterogenous findings in an analysis of 26 articles. The PACU framework includes, among others, considerations about agent design which distinguishes between the global design level (e.g., human vs. non-human, and static vs. animated), the medium design level (e.g., partial vs. full display, audio output, and language style), and the detail design level (e.g., age, gender, and clothing of the pedagogical agent).

Further reviews (Martha and Santoso, 2019; Sikström et al., 2022) and meta-analyses (Castro-Alonso et al., 2021) concluded that the inclusion of pedagogical agents facilitates learning with multimedia applications, even though the overall effects were rather small. However, in the studies analyzed by Castro-Alonso et al. (2021), there was significant heterogeneity regarding the observed effect sizes and moderation analyses revealed that effects of pedagogical agents were mostly moderated by characteristics of the studies (e.g., the design of the control group without the agent and type of randomization), but hardly by the agents' design characteristics such as appearance (2D or 3D), non-verbal communication (gestures, gaze, and facial expressions), motion (static or animated), or voice (human or synthesized). On the one hand, this could be attributed to power problems because there were only few studies in some of the relevant categories for the moderation analyses, but it could on the other hand also hint toward the need to investigate other agent characteristics as potential moderators. In contrast to Castro-Alonso et al. (2021) and Dai et al. (2022) conclude that the success of pedagogical agents highly depends on the agents' design, especially regarding the level of embodiment by means of gestures and facial expressions. However, the agent's clothing was not included in the analysis. Further, Tao et al. (2022) focused on studies which covered different variations of agent design such as human likeness (from cartoon-like to photo-realistic) or gesturing. They concluded that an agent's degree of human likeness and attractiveness could influence learning, but the results were mixed considering the question which agents are preferable. Again, the effects of the agents' clothing were not considered. Finally, Liew and Tan (2021) focused their systematic review on the agent's design, especially concerning cues for an agent's expertise and emphasized the agent's appearance, including clothing as one possible cue for expertise. With regard to clothing, they particularly emphasize that expertise cues could be conveyed if the pedagogical agent's clothing reflects the social expectation of the clothing of an expert in their domain.

Since the appearance of a pedagogical agent may be linked to its perceived expertise, it might play an important role how learners typically imagine the appearance of a particular occupational group (Küster et al., 2019; Liew and Tan, 2021). In this regard, Veletsianos (2010) observed that the matching of an agent's visual appearance to the teaching context affects learning outcomes. In a 2 × 2 design, they varied whether the agent was portrayed as a scientist or an artist by varying the characteristics of the face (neat short hairstyle without beard vs. mohawk with chin beard) and whether the agent talked about punk rock or nanotechnology. They found that the participants rated the artist agent as more knowledgeable than the scientist agent in the punk rock condition. However, both agents were rated as almost equally knowledgeable in the scientific nanotechnology domain. This finding may be due to stronger stereotypes regarding the facial appearance of a punk rocker compared to the facial appearance of a scientist, thus facilitating the visual manipulation in the punk rock domain. Further, Liew et al. (2013) also emphasized that for designing an expert-like agent, it is crucial how learners imagine experts in a field. Their manipulation of an expert-like agent vs. a peer-like agent was established by age and voice (calm vs. authoritarian). Although no significant effects

of the agents' expertise on learning outcomes were observed, female learners rated the expert-like agent as more trustworthy (Liew et al., 2013). Moreover, matching the appearance of a pedagogical agent with the associations triggered by the learning material, resulted in superior learning in an experiment (2 × 2) in that the age of the agent (young vs. old) was crossed with content relating to these age categories (relating to young vs. old age) (Beege et al., 2017).

In the current study, we investigated the effects of clothing as an attribute of pedagogical agents (detail design level, see Heidig and Clarebout, 2011). In particular, clothing which is typical for an expert in a field could influence the perceived expertise of a pedagogical agent (Baylor, 2011; Küster et al., 2019). Therefore, in the following section, we summarize research that focuses on the effects of clothing of both humans (e.g. Johnson et al., 2014; Beege et al., 2019) and pedagogical agents (Schmidt et al., 2019; Petersen et al., 2021).

1.3. Effects of clothing

In general, clothing is considered to be a form of communication that reveals information about the wearer (Damhorst, 1990). Following symbolic interactionism (Blumer, 1969), our behavior toward a person is influenced by their clothing and the meaning we attach to it.

In particular, clothing that is typically worn by members of a certain professional group (e.g., soldiers, policemen, doctors, postmen, judges, or priests) is often accompanied by an attribution of personal expertise (Kodžoman, 2019). Therefore, not only the perception of the competence often varies with a persons' clothing (Gurney et al., 2017), but also the trust invested in them. For instance, Hartmans (2013) investigated the influence of a doctor's clothing on patients' trust and their perception of medical expertise. After viewing images of six models (three male and three female) in three age groups (25–35 vs. 35–50 vs. >50), each wearing five different outfits (leisure clothing vs. casual vs. semiformal vs. formal vs. professional), participants completed a questionnaire which revealed that patients trusted most in doctors wearing a white coat. Thus, the study shows that patients prefer doctors dressed in professional clothing.

In an educational context, the quality of a presentation given by formally dressed students was rated higher than that of casually dressed students (Gurung et al., 2014). Comparably, formally dressed teachers were perceived as knowledgeable, well prepared, and organized (Lukavsky et al., 1995), but less approachable than informally dressed teachers (Butler and Roesel, 1989). Further, Beege et al. (2019) investigated whether the clothing style (professional vs. non-professional) and the addressing style (frontal vs. lateral) of an instructor affect learning. In two experiments, the authors observed that professional clothing improved retention if the instructor in the video directly looked at the camera (frontal addressing style) (Beege et al., 2019). A more recent study by Beege et al. (2022) moreover emphasizes that the effects of the instructors' professional appearance on learning and para-social interactions are enhanced when the instructor also communicates professionally.

Only few studies investigated the effect of pedagogical agents' clothing. For example, Küster et al. (2019) observed that both a nurse outfit and a military outfit were associated with more perceived competence of an agent compared to a casual outfit. However, they used static pictures of the agents and did not include any measures of learning outcomes. In contrast, in a museum context, Schmidt et al. (2019) found positive effects on learning and presence, if an agent was represented as a museum guide rather than an astronaut. However, Petersen et al. (2021) observed heterogeneous results regarding the effects of the thematically appropriate agent design on learning in a virtual museum. Comparing an agent dressed as museum guide with the same agent displaying only black mesh and a control group without an agent, they found that all the investigated pedagogical agents seemed to hinder the acquisition of factual knowledge. In contrast, regarding conceptual knowledge, thematically appropriate agents had a positive effect on learning, but only if the agent did not show realistic features (i.e., no gestures, very limited idle animations, and no lip-syncing), and was presented in a rather static way.

In summary, research on the effects of clothing on learning outcomes is not yet conclusive and it is yet an open question how a pedagogical agent's clothing interacts with other design characteristics of the learning environment. In this regard, we take a closer look at the effects of setting because setting may be another characteristic that influences learners' perceptions of a pedagogical agent's expertise.

1.4. Effects of setting

There is an ongoing debate on whether authentic settings comparable to professional contexts should be used in education and how this could be implemented (Anker-Hansen and Andreé, 2019). To this end, several studies observed that more authentic settings such as research facilities for science topics, positively influenced students' perceptions of relevance and interest (Betz, 2018; Schüttler et al., 2021).

However, whether the result that an authentic real-world setting improves perceived interest generalizes to better learning outcomes when authentic settings are implemented in learning videos is up to debate. Importantly, this question arises for both photo-realistic videos as well as for digitally animated videos (Schroeder and Adesope, 2012). For photo-realistic videos, the setting is defined as the physical environment in which the learning video was shot (see Merkt et al., 2020). Applying this definition to digitally animated videos, we define setting as the digitally designed background of the video.

To explain possible effects of setting on learning, Merkt et al. (2020) proposed that the setting of the learning video may affect the evaluation of an instructor's expertise. The learning videos in the study by Merkt et al. (2020) covered floral topics and were either shot in a botanical garden as an authentic setting or in front of a white wall as a neutral setting. Across two experiments, there were no consistent setting effects for learning outcomes, so that the authors come to the cautious conclusion that the effects of setting on learning outcomes may be negligible. Further, contrary to the authors' initial assumptions in both experiments, there were no effects of setting on the perceived expertise of the instructor. However, Merkt et al. (2020) discussed whether including an inappropriate setting instead of a neutral setting (white wall) as

a control condition may result in setting effects and different expertise judgments. Further, it is possible that expertise judgments regarding a virtual pedagogical agent may be more susceptible to external cues because virtual pedagogical agents may provide less cues that learners may use to judge expertise.

In summary, previous research did not find consistent effects of setting on learning outcomes and perceived expertise. The current experiment will therefore address the effect of an appropriate setting compared to an inappropriate setting and additionally address potential interactions with a virtual pedagogical agent's clothing.

1.5. The present research – hypotheses

Both for photo-realistic and for animated videos, producers of instructional materials are confronted with the questions what instructors/pedagogical agents should wear and in which setting the instructional materials should be produced. Based on the presented theoretical framework, it becomes evident that both of these design characteristics may affect the learning process by means of inducing different perceptions of expertise, whereas empirical evidence in favor of these assumptions is mostly mixed. In particular, previous research on clothing does imply that thematically appropriate clothing may be beneficial for learning (Beege et al., 2019; Schmidt et al., 2019), whereas the effects of a thematically appropriate setting are considered to be rather negligible (Merkt et al., 2020). Because potential effects of clothing and setting may be driven by the same underlying mechanism (i.e., perceived expertise), we decided to investigate both of these design factors in the same experiment, employing a 2 × 2-factorial between subject design. Given the lack of research combining these two factors, we decided to treat the interaction of the two factors as an open research question because there is no clear theoretical indication about the potential interplay of these two factors. On the one hand, it may be sufficient to include one cue for instructor expertise, on the other hand, it is also conceivable that the different sources for expertise amplify each other.

Building on these considerations, we preregistered the following hypotheses on OSF.

> H1: The video including the pedagogical agent wearing thematically appropriate clothing is rated as more professional than the video including the pedagogical agent wearing thematically inappropriate clothing.

> H2: The pedagogical agent wearing thematically appropriate clothing is considered to have a higher expertise than the pedagogical agent wearing thematically inappropriate clothing.

> H3: The pedagogical agent wearing thematically appropriate clothing results in better learning outcomes than the pedagogical agent wearing thematically inappropriate clothing.

Based on previous mixed evidence on the effects of setting on learning outcomes (see Merkt et al., 2020), we refrain from formulating hypotheses regarding a main effect of setting. Further, the study explores the interaction of clothing and setting.

2. Materials and methods

2.1. Participants and design

Overall, 231 participants completed the experiment. Thirty-two participants were excluded due to pre-defined exclusion criteria, namely technical difficulties during the presentation of the video (1 participant), use of search engines such as Google (7 participants), a native language other than German (2 participants), or taking notes during the learning video (21 participants). This resulted in a final sample of 200 German-speaking participants (131 male, 68 female, and 1 diverse), recruited using Prolific. The participants were randomly assigned to one of the four conditions (2 × 2 design) and received 4.38 Pounds for their participation. Their mean age was 29.52 years (SD = 8.73). Even though a power analysis revealed that 125 participants would have been sufficient to detect a medium effect ($\eta_p^2 = 0.06$) with a predefined alpha level of 0.05 and power of 0.80, we pre-registered to collect data of 200 participants to increase power and identify potential interactions.

Participants were randomly assigned to one of four conditions resulting from a 2 × 2 factorial between-subjects design with the clothing of the pedagogical agent (thematically appropriate or thematically inappropriate clothing) and the setting of the animated video (thematically appropriate or thematically inappropriate setting) as the independent variables. Learning outcomes and different evaluation questions were collected as dependent variables.

2.2. Learning materials

The learning material consisted of an animated video (duration: 19:30 min) that was presented to the participants in one of four different versions, depending on the experimental condition they were assigned to (see **Figure 1**). In the video, a pedagogical agent presented learning content on craft topics such as wood and its applications, connections, weather protection, sawing, glues, tools, locks, and corrosion. The information was partly visualized in a PowerPoint presentation that was superimposed over the background on the right-hand side of the agent. The presentation appeared only occasionally to visually highlight important content, whereas the setting was fully visible throughout the remainder of the video. The pedagogical agent had a male human voice and could be seen in its entirety (full-body representation) displaying conversational gestures (see Davis, 2018) and mimic. Neither the facial expressions nor the gestures did have any deictic functions. The pedagogical agent was created with Character Creator 3.3 and animated with iClone 7.8.

As appropriate clothing for a craft theme, the pedagogical agent wore brown dungarees. The design of this clothing was based on the recommendations for workwear of the Employer's Liability Insurance Association for Wood and Metal in Germany (Berufsgenossenschaft Holz und Metall [BGHM], 2019) and can thus be considered prototypical for the cultural context in which

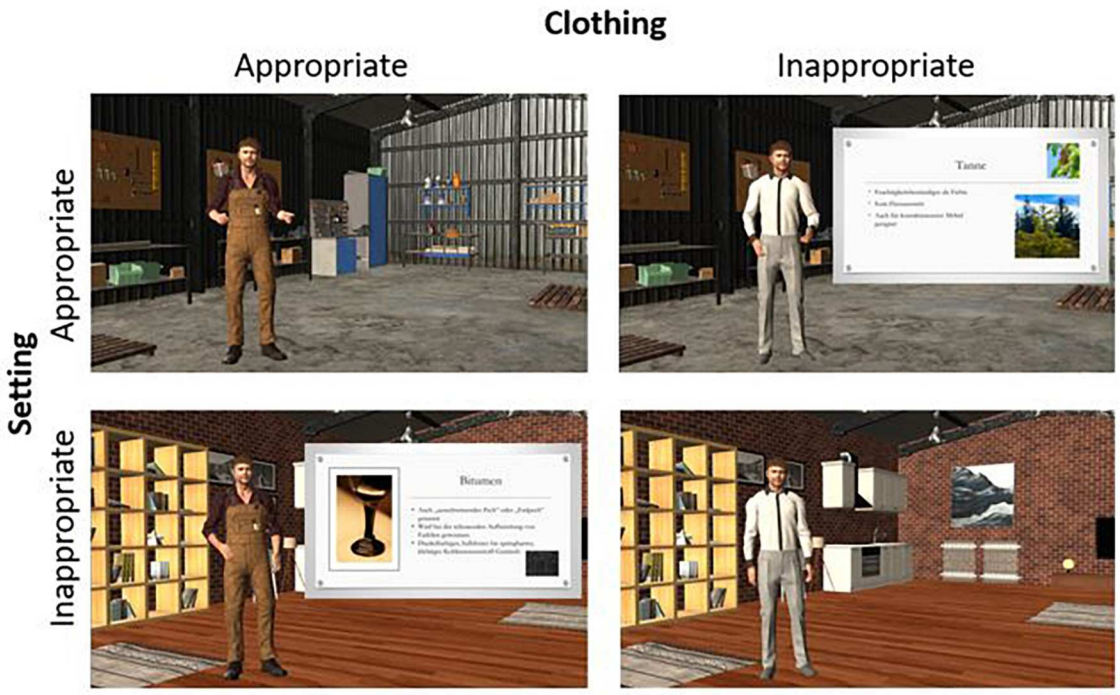


FIGURE 1
Conditions of the learning video displaying the variation of setting and clothing. To illustrate the insertion of the presentation, one version with and one without the presentation was selected for each setting.

the study was conducted. For example, according to a study on woodworking injuries, 68% of professional craftsmen wore safety clothing similar to the equipment described in the above mentioned handout (Loisel et al., 2014). In contrast, in the inappropriate clothing condition, the pedagogical agent wore a white buttoned shirt with neat gray trousers. Since the visual appearance of our setting should appear as if crafts could be carried out without difficulty, the pedagogical agent was placed in a classic workshop setting with a gray concrete floor, tool walls, a workbench, pallets, and garage-like walls in the appropriate setting condition. In contrast, for the inappropriate setting condition, the pedagogical agent was placed in a living room environment with a wooden laminar floor, shelves, paintings, a kitchenette in the background, and small rugs on the floor, where the practical execution of craft activities would be more difficult. The arrangement, as well as the number and size of the objects in the background of both settings were chosen to match as closely as possible. It was ensured that the videos did not differ regarding any other characteristics than clothing and setting, especially because facial expressions, gestures, and speed of delivery could be kept perfectly constant due to the digital nature of the videos.

2.3. Measures and instruments

2.3.1. Prior knowledge and interest

Prior knowledge and interest in the learning topics were each measured with 5 items on a 7-point Likert scale in self-assessment (cf. Harp and Mayer, 1997; Kühl et al., 2019; Sondermann and Merkt, 2023), ranging from 1 (almost no prior knowledge/interest at all) to 7 (a lot of prior knowledge/very great interest). The topics covered in these questions were: Areas of use for different types of wood, wood connection, weather protection, tools, and corrosion. Participants' scores for prior interest and prior knowledge were determined by averaging their ratings for each of the topics. Cronbach's alphas for the prior knowledge items and the interest items were 0.92 and 0.93, respectively.

2.3.2. Evaluation of the learning video

Following the research of Merkt et al. (2020), participants were asked to rate the video on multiple 7-point Likert scales from 1 (lowest score) to 7 (highest score) covering different video characteristics. These characteristics included expertise of the instructor ("How much of an expert was the instructor?"), professionalism ("How professional did the video appear to be?"), quality of explanations ("How would you rate the quality of the explanations?"), and distraction by the setting ("How distracting was the setting of the video?"). As a manipulation check, we also measured the fit between the agent's clothing ("How would you rate the fit between the teaching theme and the avatar's[1] clothing?") and the setting ("How would you rate the fit between the teaching topic and the setting of the video?") to the topics presented. Further, we measured the invested mental effort ("How much effort did you invest in following the contents of the video?"; see Paas, 1992).

1 Since our study took place in Germany, where the term "pedagogical agent" is not very common, we decided to refer to our pedagogical agent as avatar in the evaluation questions to avoid misunderstandings among the participants.

Additionally, we measured credibility of the pedagogical agent ("How credible do you perceive the avatar to be?"), naturalness of gestures and facial expressions of the pedagogical agent ("How natural did you find the gestures of the avatar?" and "How natural did you find the facial expressions of the avatar?"), perceived appropriateness of the video length ("How appropriate did you find the length of the video?"), the perceived personal addressing ("It seemed like the lecturer was speaking directly to me"; see Beege et al., 2019) and the perceived difficulty ("How difficult was the content of the video?").

2.3.3. Knowledge test

The learning outcome was assessed with 20 multiple choice questions with four possible answers, one of which was correct. For each question, participants received one point for selecting the correct answer, resulting in a maximum score of 20 points. Cronbach's alpha for the knowledge items was 0.73. The items included questions such as "What is the average humidity of free-standing wood without weather protection?" (answer options: 18 vs. 12 vs. 22 vs. 25%, correct answer: 18%).

To assess whether the questions in the knowledge test could be answered based on general knowledge, we piloted the items in a sample of 20 participants (9 female and 11 male). The pilot participants' mean age was 29.15 years (SD = 9.07). The pilot study was run on LimeSurvey and the participants were recruited via Prolific, an online survey platform with a large number of registered participants. They gave informed consent and were instructed to answer the questions solely based on their knowledge without using search engines. Afterward they filled in a knowledge test consisting of 39 multiple choice questions without any previous access to learning materials. From this pool of items from the pilot study, we then selected 20 items for the current study that showed adequate moderate difficulty. The mean score for those 20 items of the knowledge test in the pilot study answered without learning materials was M = 7.65 (SD = 2.03). Thus, whereas a small proportion of the questions could be answered based on general knowledge, the test cannot be considered too easy and provides room for improvement after watching the learning video.

2.4. Procedure

The experiment was run online using LimeSurvey with participants recruited via Prolific. After participants agreed to the consent form, they were randomly assigned to one of four experimental conditions. Afterward, a picture and sound test was administered to ensure that the learning video was correctly presented with picture and sound. Next, prior knowledge and general interest in the learning content was recorded. One of the four versions of the learning video was presented subsequently, according to the experimental condition the participants were randomly assigned to. Following the video, participants were asked to evaluate the video before they filled in the knowledge test. After the knowledge test, the variables regarding the exclusion criteria (technical difficulties during the presentation of the video, use of search engines, a native language other than German, or taking notes) were assessed with an explicit request for honesty and an additional remark that this information was important for assessing data quality but would not affect payment. Finally, demographic data were queried, and the debriefing took place. The procedure was approved by the Local Ethics Committee of the German Institute for Adult Education.

3. Results

Data were analyzed with 2 × 2 ANOVAs with the independent variables Clothing (thematically appropriate vs. inappropriate) and Setting (thematically appropriate vs. inappropriate) and the respective outcomes as dependent variables. Further, to test whether the video was suitable for knowledge acquisition, a one sample t-test was used to compare the mean score that participants had achieved in the pilot study of the knowledge test (i.e., 7.65 points), in which the items were tested without learning materials, to the mean score achieved in this experiment over all conditions.

The data and the analysis script are available on OSF. Descriptive data are presented in **Table 1**.

3.1. Manipulation check

Regarding the perceived appropriateness of the clothing, the thematically appropriate clothing was rated to be more suitable (M = 5.94, SD = 0.97) for the topic than the thematically inappropriate clothing (M = 3.78, SD = 1.55), $F(1,196)$ = 138.89, $p < 0.001$, $\eta_p^2 = 0.41$. There was neither a main effect of the Setting, $F < 1$; nor an interaction Clothing × Setting, $F < 1$.

Regarding the appropriateness of the setting, the thematically appropriate setting was rated as more appropriate (M = 5.36, SD = 1.38) than the thematically less appropriate setting (M = 4.00; SD = 1.62), $F(1,196)$ = 41.10, $p < 0.001$, $\eta_p^2 = 0.17$. There was also neither a main effect of Clothing, $F(1,196)$ = 3.56, p = 0.061, $\eta_p^2 = 0.02$; nor an interaction Clothing × Setting, $F < 1$.

3.2. Prior knowledge, interest, and demographics

With regard to self-reported prior knowledge (M = 2.39; SD = 1.15), there were no main effects of Clothing, $F(1,196)$ = 2.56, p = *0.111*, $\eta_p^2 = 0.01$, and Setting, $F < 1$, and no interaction Setting × Clothing, $F < 1$. Likewise, for the self-reported interest (M = 3.33; SD = 1.49), there were no main effects of Clothing, $F(1,196)$ = 1.07, p = 0.302, $\eta_p^2 = 0.01$, and Setting, $F < 1$, as well as no interaction Clothing × Setting, $F < 1$. Thus, we can assume that the groups were comparable before the experiment.

Further, neither the proportion of gender X^2 (6, N = 200) = 7.50, p = 0.277, nor the proportion of highest educational qualifications, X^2 (15, N = 200) = 15.46, p = 0.419, differed between the groups. With regard to age, there were no main effects of Clothing, $F < 1$, and Setting, $F < 1$, and no interaction Clothing × Setting, $F(1,196)$ = 1.49, p = 0.224, $\eta_p^2 = 0.01$. Thus, we can assume that the groups were comparable regarding demographic criteria.

TABLE 1 Overall means and standard deviations.

Variable	Appropriate clothing			Inappropriate clothing			Overall		Overall
	Appropriate setting	Inappropriate setting	Overall	Appropriate setting	Inappropriate setting	Overall	Appropriate setting	Inappropriate setting	
Prior interest	3.29 (1.49)	3.15 (1.29)	3.22 (1.39)	3.43 (1.61)	3.45 (1.60)	3.44 (1.60)	3.36 (1.54)	3.30 (1.46)	3.33 (1.49)
Prior knowledge	2.34 (1.02)	2.19 (1.20)	2.26 (1.11)	2.49 (1.06)	2.56 (1.29)	2.52 (1.18)	2.41 (1.04)	2.37 (1.25)	2.39 (1.15)
Knowledge test	13.14 (4.06)	14.58 (3.50)	13.86 (3.84)	13.96 (3.75)	13.34 (3.23)	13.65 (3.49)	13.55 (3.91)	13.96 (3.41)	13.76 (3.67)
Professionalism	3.80 (1.48)	4.08 (1.62)	3.94 (1.56)	3.78 (1.43)	3.66 (1.72)	3.72 (1.58)	3.79 81.45)	3.87 (1.68)	3.83 (1.57)
Expertise	4.42 (1.72)	4.08 (1.87)	4.25 (1.79)	3.96 (1.76)	3.72 (1.98)	3.84 (1.87)	4.19 (1.75)	3.90 (1.92)	4.04 (1.84)
Quality	5.02 (1.13)	5.30 (1.15)	5.16 (1.19)	5.10 (1.13)	4.96 (1.32)	5.03 (1.23)	5.06 (1.17)	5.13 (1.24)	5.10 (1.21)
Naturalness – gestures	2.44 (1.37)	2.72 (1.39)	2.58 (1.38)	2.46 (1.42)	2.46 (1.49)	2.56 (1.45)	2.45 (1.39)	2.58 (1.44)	2.52 (1.41)
Naturalness – mimic	2.34 (1.17)	2.72 (1.32)	2.53 (1.26)	2.44 (1.45)	2.28 (1.46)	2.36 (1.45)	2.39 (1.31)	2.50 (1.40)	2.45 (1.36)
Difficulty	3.64 (1.36)	3.34 (1.47)	3.49 (1.42)	3.46 (1.41)	3.28 (1.43)	3.37 (1.42)	3.55 (1.39)	3.31 (1.44)	3.43 (1.42)
Effort invested	4.70 (1.54)	4.54 (1.58)	4.62 (1.56)	4.78 (1.94)	4.74 (1.65)	4.76 (1.79)	4.74 (1.74)	4.64 (1.61)	4.69 (1.68)
Credibility	4.52 (1.75)	4.58 (1.86)	4.55 (1.80)	3.78 (1.75)	3.92 (1.94)	3.85 (1.84)	4.15 (1.78)	4.25 (1.92)	4.20 (1.85)
Thematical fit – clothing	6.06 (1.04)	5.82 (0.89)	5.94 (0.97)	3.82 (1.64)	3.74 (1.45)	3.78 (1.55)	4.94 (1.77)	4.78 (1.59)	4.86 (1.68)
Thematical fit – setting	5.56 (1.16)	4.2 (1.55)	4.88 (1.52)	5.16 (1.55)	3.8 (1.68)	4.48 (1.74)	5.36 (1.38)	4.00 (1.62)	4.68 (1.65)
Distraction	2.58 (1.25)	2.96 (1.64)	2.77 (1.46)	2.76 (1.66)	3.32 (1.66)	3.04 (1.68)	2.67 (1.46)	3.14 (1.65)	2.91 (1.57)
Length	3.86 (1.60)	3.68 (1.51)	3.77 (1.56)	3.70 (1.71)	3.78 (1.60)	3.74 (1.66)	3.78 (1.65)	3.73 (1.56)	3.76 (1.60)
Personal address	3.24 (1.57)	3.12 (1.63)	3.18 (1.60)	3.08 (1.77)	2.94 (1.78)	3.01 (1.77)	3.16 (1.67)	3.03 (1.70)	3.09 (1.68)
N	50	50	100	50	50	100	100	100	200

10.3389/fpsyg.2023.1205338

3.3. Perceived professionalism and expertise

Regarding the first two hypotheses, the analyses revealed no significant differences between the conditions. For perceived professionalism of the video (M = 3.83, SD = 1.57), there were no main effects of Clothing and Setting and no significant interaction Clothing × Setting, all $F < 1$. Also, concerning perceived expertise of the agent (M = 4.04; SD = 1.84), there were no significant main effects of Clothing, $F(1,196) = 2.49$, $p = 0.116$, $\eta_p^2 = 0.012$, and Setting, $F(1,196) = 1.25$, $p = 0.266$, $\eta_p^2 = 0.01$, and no significant interaction Clothing × Setting, $F < 1$.

3.4. Learning outcomes

For learning outcomes, there were no significant main effects of either Clothing or Setting, both $F < 1$. However, there was an interaction Clothing × Setting, $F(1,196) = 3.98$, $p = 0.047$, $\eta_p^2 = 0.02$. Following up on this interaction, Bonferroni-adjusted *post-hoc* comparisons revealed that there was no significant effect of setting for the inappropriate clothing $F < 1$, whereas in the appropriate clothing condition, the inappropriate setting (M = 14.58, SD = 3.50) resulted in slightly better learning outcomes than the appropriate setting (M = 13.14, SD = 4.06), $F(1,196) = 3.89$, $p = 0.050$, $\eta_p^2 = 0.02$. When the interaction was resolved the other way, the Bonferroni-adjusted *post-hoc* tests showed no effect of clothing, neither for the appropriate setting, $F(1,196) = 1.26$, $p = 0.263$, $\eta_p^2 = 0.006$, nor for the inappropriate setting, $F(1,196) = 2.89$, $p = 0.091$, $\eta_p^2 = 0.015$.

Further, results of a one sample t-test showed that participants in all conditions (M = 13.76; SD = 3.67) learned from the video because their mean score was higher than the mean score of participants in the pilot study (i.e., M = 7.65) that had answered the same items without the learning material, $t(199) = 23.55$, $p < 0.001$.

3.5. Further evaluation variables

For distraction, there was a significant main effect of Setting, revealing that the inappropriate setting (M = 3.14, SD = 1.65) was rated to be more distracting than the appropriate setting (M = 2.67, SD = 1.46), $F(1,196) = 4.53$, $p = 0.035$, $\eta_p^2 = 0.02$. There was no main effect of Clothing, $F(1,196) = 1.49$, $p = 0.223$, $\eta_p^2 = 0.01$; and no interaction Clothing × Setting, $F < 1$.

For credibility, there was a main effect of Clothing, showing that the appropriately dressed pedagogical agent was rated as more credible (M = 4.55, SD = 1.80) than the thematically inappropriately dressed agent (M = 3.85, SD = 1.84), $F(1,196) = 7.31$, $p = 0.007$, $\eta_p^2 = 0.04$. There was no main effect of Setting, $F < 1$, and no interaction Clothing × Setting, $F < 1$.

There were no main effects or interactions of Clothing and Setting in terms of perceived difficulty, facial expressions of the pedagogical agent, naturalness of gestures, mental effort invested, quality of explanations, perceived appropriateness of the video length and perceived personal addressing, all $F < 1.44$, all $p > 0.159$.

4. Discussion

In this experiment, we investigated the effects of a pedagogical agent's clothing and the setting of an animated video on learning outcomes and participants' subjective evaluations such as perceived agent expertise and professionalism of the video. Overall, the experiment did not provide any strong evidence for systematic effects of a pedagogical agent's clothing and the setting of an animated learning video. Contrary to our expectations, the thematically appropriate clothing did not result in more perceived professionalism of the video (Hypothesis 1) or perceived expertise of the agent (Hypothesis 2) than the thematically inappropriate clothing, even though the manipulation check clearly showed that the thematically appropriate clothing was perceived as more appropriate and that the appropriately dressed agent was perceived as more credible than the inappropriately dressed agent. In this regard, it should be mentioned that even though the inappropriate clothing resulted in lower appropriateness and credibility judgments, the inappropriate (i.e., formal) clothing was still judged to be moderately appropriate (M = 3.78 on a scale ranging from 1 to 7) and the agent was judged to be moderately credible (M = 3.85 on a scale ranging from 1 to 7) in the formal clothes. A possible explanation for this could be that both thematically appropriate clothing and formal clothing exude expertise to the same extent, as formal clothing may be associated with expertise in different fields. Whereas we used formal clothing because we considered it to be inappropriate in a handicraft context, future studies may want to contrast thematically appropriate clothing with clothing that is usually strongly associated with other professions. In the context of a handicraft topic, white coats or police uniforms could be examples for highly inappropriate clothing since they might be strongly associated with other occupations.

Further, contrary to Hypothesis 3, appropriate clothing did not lead to better learning outcomes than the formal clothing in the inappropriate condition.

Regarding the open research question concerning the effect of setting on the learning outcomes, we did not find any setting effects, corroborating cautious conclusions from previous research that there may be no overall setting effects (see Merkt et al., 2020). However, similar to the issues discussed regarding our manipulation of clothing, the living room setting in the inappropriate setting condition was rated to be moderately appropriate (score of 4.00 on a scale ranging from 1 to 7), so that we may have succeeded in manipulating the level of appropriateness but did not manage to provide learners with an inappropriate setting.

However, an interaction between clothing and setting may shed light on some potential boundary conditions for setting effects. In particular, the inappropriate setting resulted in better learning outcomes than the appropriate setting when the pedagogical agent wore appropriate clothing, whereas there was no setting effect when the pedagogical agent wore inappropriate clothing. The result that the appropriate setting did not lead to better learning regardless of the clothing could be explained with the expectancy violations theory (Burgoon, 1993; Dunbar and Segrin, 2012). The appropriate setting, especially combined with the appropriate clothing of the agent may have created the false expectation that something

practical would be demonstrated on the machines depicted in the background of the authentic setting. These unmet expectations may have been detrimental to the learning outcomes. In this experiment, the pedagogical agent did not demonstrate how to use the authentic machines because we aimed at keeping the learning content constant across the appropriate and the inappropriate setting conditions. Therefore, interactions with the machines could not be included since the machines were only present in half of the conditions. To determine whether the violation of expectations played a crucial role, participants in future experiments could be asked how they imagine an expert in the field and what they expect from the learning materials before the learning phase. Further, because this unexpected effect is very small ($\eta_p^2 = 0.02$), it should be replicated before jumping to premature conclusions.

Finally, the rather long duration of the video (19:30 min) may have leveled out potential clothing and setting effects because these may wear off over time, as learners may pay more attention to clothing and setting at the beginning of the video. In particular, Guo et al. (2014) reported that general attention while watching videos significantly decreases after 6 min. Since clothing and setting did not change throughout the video, it could have faded into the background of attention while participants watched the video for approximately 20 min. Testing video length as a potential boundary condition for clothing and setting effects may thus be an interesting pathway for future research. These studies could additionally use process data such as eye-tracking to determine whether the attention wears off over time.

4.1. Limitations

The results are subject to some limitations. First, we tested the effect of clothing and setting with only one specific exemplar for appropriate and inappropriate manifestations. However, looking at clothing, craft clothing can include a wide range of clothing, especially as many sub-disciplines have their own specific clothing and safety equipment (such as painters and varnishers). Thus, the generalizability of these findings should be tested for more sub-disciplines in future experiments. Further, the subject areas should be extended to other occupational groups, as many other professions also have distinctive professional clothing that could be strongly associated with expertise (e.g., doctors, nurses, cooks, security guards, flight attendants, and pilots). Similar limitations should be considered for the setting. The range of settings potentially perceived as appropriate could be narrower or wider depending on the occupational group. Craft topics have a rather wide range, as for example, car repair shops look very different from carpentry shops. Regarding these factors, future research questions in this area should be chosen accordingly.

Further, it should be noted that the interpretation of our findings regarding specific pieces of clothing and the setting as depicted in **Figure 1** is limited to the cultural background of the study (i.e., Germany). In particular, the appropriateness of different pieces of clothing and different types of setting for different purposes may vary with culture. However, it should also be noted that the investigated underlying principle of thematical appropriateness should be more generalizable across cultures if researchers from different cultures carefully select pieces of clothing and a setting that is considered appropriate for a specific topic in their culture. Further, the generalizability of these findings across cultures could also be tested to see whether there are different effects of the same learning materials in different cultures.

Whereas the use of different types of clothing to evoke different perceptions of expertise could be considered a problematic case of stereotyping, we want to stress that this research does not endorse the use of stereotyping to manipulate learners. Therefore, this study did not use exaggerated stereotypical clothing, but rather prototypical clothing which is usually worn by persons of a specific occupation (Berufsgenossenschaft Holz und Metall [BGHM], 2019). Because some types of clothing may be closely linked to specific occupations (see Kodžoman, 2019), we were interested whether learners use these associations between clothing and specific occupations as a heuristic to determine a pedagogical agents' expertise. If such effects had occurred, it would have been an interesting follow up question how to handle and prevent such effects.

Further, since the elaboration likelihood model assumes that heuristics are often used as a shortcut instead of more elaborated cognitive processes (Petty and Cacioppo, 1986), it is feasible to assume that learners may rely on heuristics more heavily when their cognitive load is high. However, since it was not the aim of our study to investigate whether cognitive load influenced the usage of heuristics for evaluating the expertise of a pedagogical agent, we did not measure cognitive load specifically, but decided to use the item by Paas (1992) to gain insights on the effects of mental effort. Since cognitive load and mental effort are distinct concepts that affect learning outcomes in different ways (see Schnaubert and Schneider, 2022), the question whether heuristics play an increased role under high cognitive load cannot be answered with this experiment and should therefore be addressed in further research that measures cognitive load using more nuanced instruments that assess different facets of cognitive load (Klepsch et al., 2017; Krieglstein et al., 2023).

5. Conclusion

This experiment tried to shed light on the question whether it is necessary to design and animate thematically appropriate clothing for pedagogical agents and add an appropriate setting to an animated video to increase learning outcomes and perceived expertise of the agent as well as perceived professionalism of the learning video. The results imply that in case of a handicraft topic, designers of animated videos may have some degrees of freedom when selecting the clothing of the agent and the setting of the video because participants in all conditions acquired knowledge and the influence of clothing and setting on learning outcomes and perceived expertise was negligible. However, it should be noted that none of the clothing and setting in our study was perceived to be inappropriate for the learning materials. Thus, the use of a highly inappropriate clothing and setting may still constitute a boundary condition for clothing and setting effects.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found below: OSF – https://osf.io/qm4pv/?view_only=1715aaa02cb74e588b591882b08a3bd3.

Ethics statement

The studies involving humans were approved by the Local Ethics Committee of the German Institute for Adult Education – Leibniz Center for Lifelong Learning e.V. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

DD: writing—original draft, methodology, editing, conceptualization, and data analysis. MM: writing—reviewing, conceptualization, and supervision. Both authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anker-Hansen, J., and Andreé, M. (2019). In pursuit of authenticity in science education. *Nordic Stud. Sci. Educ.* 15, 54–66. doi: 10.5617/nordina.4723

Apoki, U. C., Ali Hussein, A. M., Al-Chalabi, H. K. M., Badica, C., and Mocanu, M. L. (2022). The role of pedagogical agents in personalised adaptive learning: A review. *Sustainability* 14:6442. doi: 10.3390/su14116442

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Baylor, A. L. (2011). The design of motivational agents and avatars. *Educ. Technol. Res. Dev.* 59, 291–300. doi: 10.1007/s11423-011-9196-3

Beege, M., Krieglstein, F., and Arnold, C. (2022). How instructors influence learning with instructional videos - The importance of professional appearance and communication. *Comput. Educ.* 185:104531. doi: 10.1016/j.compedu.2022.104531

Beege, M., Nebel, S., Schneider, S., and Rey, G. D. (2019). Social entities in educational videos: Combining the effects of addressing and professionalism. *Comput. Hum. Behav.* 93, 40–52. doi: 10.1016/j.chb.2018.11.051

Beege, M., Schneider, S., Nebel, S., Mittangk, J., and Rey, G. D. (2017). Ageism – Age coherence within learning material fosters learning. *Comput. Hum. Behav.* 75, 510–519. doi: 10.1016/j.chb.2017.05.042

Berufsgenossenschaft Holz und Metall [BGHM] (2019). *DGUV Regel 109-606 Branche Tischler- und Schreinerhandwerk* (Issue April). Available online at: https://www.bghm.de/fileadmin/user_upload/Arbeitsschuetzer/Gesetze_Vorschriften/Regeln/109-606.pdf (accessed September 15, 2023).

Betz, A. (2018). The influence of the learning environment on learners' (perceived) authenticity of science communication and on their situational interest. *Unterrichtswissenschaft* 46, 261–278. doi: 10.1007/s42010-018-0021-0

Bian, Y. (2022). Motivation effect of animated pedagogical agent's personality and feedback strategy types on learning in virtual training environment. *Virtual Reality Intellig. Hardw.* 4, 153–172. doi: 10.1016/j.vrih.2021.11.001

Blumer, H. (1969). *Symbolic interactionism; perspective and method*. Englewood Cliffs, NJ: Prentice-Hall.

Burgoon, J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *J. Lang. Soc. Psychol.* 12, 30–48. doi: 10.1177/0261927X93121003

Butler, S., and Roesel, K. (1989). Research note: The influence of dress on students' perceptions of teacher characteristics. *Cloth. Textiles Res. J.* 7, 57–59. doi: 10.1177/0887302X8900700309

Castro-Alonso, J. C., Wong, R. M., Adesope, O. O., and Paas, F. (2021). Effectiveness of multimedia pedagogical agents predicted by diverse theories: A meta-analysis. *Educ. Psychol. Rev.* 33, 989–1015. doi: 10.1007/s10648-020-09587-1

Craig, S. D., Gholson, B., and Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *J. Educ. Psychol.* 94, 428–434. doi: 10.1037/0022-0663.94.2.428

Dai, L., Jung, M. M., Postma, M., and Louwerse, M. M. (2022). A systematic review of pedagogical agent research: Similarities, differences and unexplored aspects. *Comput. Educ.* 190:104607. doi: 10.1016/j.compedu.2022.104607

Damhorst, M. L. (1990). In search of a common thread: Classification of information communicated through dress. *Cloth. Textiles Res. J.* 8, 1–12. doi: 10.1177/0887302X9000800201

Davis, R. O. (2018). The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis. *Educ. Res. Rev.* 24, 193–209. doi: 10.1016/j.edurev.2018.05.002

Dunbar, N. E., and Segrin, C. (2012). Clothing and teacher credibility: An application of expectancy violations theory. *ISRN Educ.* 2012, 1–12. doi: 10.5402/2012/140517

Grivokostopoulou, F., Kovas, K., and Perikos, I. (2020). The effectiveness of embodied pedagogical agents and their impact on students learning in virtual worlds. *Appl. Sci.* 10:1739. doi: 10.3390/app10051739

Guo, P. J., Kim, J., and Rubin, R. (2014). "How video production affects student engagement: An empirical study of MOOC videos," in *Proceedings of the First ACM Conference on Learning@ Scale Conference*, (New York, NY: ACM Press), 41–50.

Gurney, D. J., Howlett, N., Pine, K., Tracey, M., and Moggridge, R. (2017). Dressing up posture: The interactive effects of posture and clothing on competency judgements. *Br. J. Psychol.* 108, 436–451. doi: 10.1111/bjop.12209

Gurung, R. A. R., Kempen, L., Klemm, K., Senn, R., and Wysocki, R. (2014). Dressed to present: Ratings of classroom presentations vary with attire. *Teach. Psychol.* 41, 349–353. doi: 10.1177/0098628314549710

Harp, S. F., and Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *J. Educ. Psychol.* 89, 92–102. doi: 10.1037/0022-0663.89.1.92

Hartmans, C. (2013). The doctor's new clothes: Professional or fashionable? *Primary Health Care Open Access* 3:145. doi: 10.4172/2167-1079.1000145

Heidig, S., and Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educ. Res. Rev.* 6, 27–54. doi: 10.1016/j.edurev.2010.07.004

Johnson, K., Lennon, S. J., and Rudd, N. (2014). Dress, body and self: Research in the social psychology of dress. *Fashion Textiles* 1:20. doi: 10.1186/s40691-014-0020-7

Klepsch, M., Schmitz, F., and Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Front. Psychol.* 8:1997. doi: 10.3389/fpsyg.2017.01997

Kodžoman, D. (2019). The psychology of clothing. *Textile Leather Rev.* 2, 90–103. doi: 10.31881/tlr.2019.22

Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., and Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure

different types of cognitive load. *Educ. Psychol. Rev.* 35:9. doi: 10.1007/s10648-023-09738-0

Kühl, T., Moersdorf, F., Römer, M., and Münzer, S. (2019). Adding emotionality to seductive details—Consequences for learning? *Appl. Cogn. Psychol.* 33, 48–61. doi: 10.1002/acp.3477

Küster, D., Krumhuber, E. G., and Hess, U. (2019). You are what you wear: Unless you moved—effects of attire and posture on person perception. *J. Nonverb. Behav.* 43, 23–38. doi: 10.1007/s10919-018-0286-3

Lange, P. G. (2018). "Informal learning on YouTube," in *The international encyclopedia of media literacy*, eds R. Hobbs, P. Mihailidis, G. Cappello, M. Ranieri, and B. Thevenin (Hoboken, NJ: John Wiley & Sons), doi: 10.1002/9781118978238.ieml0090

Lester, J. C., Barlow, S. T., Converse, S. A., Stone, B. A., Kahler, S. E., and Bhogal, R. S. (1997). "Persona effect: Affective impact of animated pedagogical agents," in *Proceedings of the Conference on human factors in computing systems*, ed. S. Pemberton (New York, NY: ACM Press), 359–366.

Li, W., Wang, F., Mayer, R., and Liu, T. (2022). Animated pedagogical agents enhance learning outcomes and brain activity during learning. *J. Comput. Assisted Learn.* 38, 621–637. doi: 10.1111/jcal.12634

Liew, T. W., and Tan, S.-M. (2021). Social cues and implications for designing expert and competent artificial agents: A systematic review. *Telemat. Inform.* 65:101721. doi: 10.1016/j.tele.2021.101721

Liew, T. W., Tan, S. M., and Jayothisa, C. (2013). The effects of peer-like and expert-like pedagogical agents on learners' agent perceptions, task-related attitudes, and learning achievement. *Educ. Technol. Soc.* 16, 275–286.

Loisel, F., Bonin, S., Jeunet, L., Pauchot, J., Tropet, Y., and Obert, L. (2014). Woodworking injuries: A comparative study of work-related and hobby-related accidents. *Chir. Main* 33, 325–329. doi: 10.1016/j.main.2014.06.003

Lukavsky, J., Butler, S., and Harden, A. J. (1995). Perceptions of an instructor: Dress and students' characteristics. *Percept. Motor Skills* 81, 231–240. doi: 10.2466/pms.1995.81.1.231

Martha, A. S. D., and Santoso, H. B. (2019). The design and impact of the pedagogical agent: A systematic literature review. *J. Educ. Online* 16, 1–15. doi: 10.9743/jeo.2019.16.1.8

Mayer, R. E. (2014). "The Cambridge handbook of multimedia learning," in *Cambridge handbooks in psychology*, 2nd Edn, ed. R. E. Mayer (Cambridge: Cambridge University Press), doi: 10.1017/CBO9781139547369

Mayer, R. E., Sobko, K., and Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *J. Educ. Psychol.* 95, 419–425. doi: 10.1037/0022-0663.95.2.419

Merkt, M., Lux, S., Hoogerheide, V., van Gog, T., and Schwan, S. (2020). A change of scenery: Does the setting of an instructional video affect learning? *J. Educ. Psychol.* 112, 1223–1283. doi: 10.1037/edu0000414

Nguyen, H. (2023). Role design considerations of conversational agents to facilitate discussion and systems thinking. *Comput. Educ.* 192:104661. doi: 10.1016/j.compedu.2022.104661

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *J. Educ. Psychol.* 84, 429–434. doi: 10.1037/0022-0663.84.4.429

Petersen, G. B., Mottelson, A., and Makransky, G. (2021). "Pedagogical agents in educational VR: An in the wild study," in *Proceedings of the conference on human factors in computing systems*, (New York, NY: Association for Computing Machinery), doi: 10.1145/3411764.3445760

Petty, R. E., and Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Adv. Exp. Soc. Psychol.* 19, 123–205. doi: 10.1016/S0065-2601(08)60214-2

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educ. Res. Rev.* 7, 216–237. doi: 10.1016/j.edurev.2012.05.003

Schmidt, S., Bruder, G., and Steinicke, F. (2019). Effects of virtual agent and object representation on experiencing exhibited artifacts. *Comput. Graphics* 83, 1–10. doi: 10.1016/j.cag.2019.06.002

Schnaubert, L., and Schneider, S. (2022). Analysing the relationship between mental load or mental effort and metacomprehension under different conditions of multimedia design. *Front. Educ.* 6:648319. doi: 10.3389/feduc.2021.648319

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., and Rey, G. D. (2022). The Cognitive-Affective-Social Theory of Learning in digital Environments (CASTLE). *Educ. Psychol. Rev.* 34, 1–38. doi: 10.1007/s10648-021-09626-5

Schroeder, N., and Adesope, O. (2012). A case for the use of pedagogical agents in online learning environments. *J. Teach. Learn. Technol.* 1, 43–47.

Schüttler, T., Watzka, B., Girwidz, R., and Ertl, B. (2021). Die Wirkung der Authentizität von Lernort und Laborgeräten auf das situationale Interesse und die Relevanzwahrnehmung beim Besuch eines naturwissenschaftlichen Schülerlabors. *Zeitschr. Didaktik Naturwissenschaften* 27, 109–125. doi: 10.1007/s40573-021-00128-z

Sikström, P., Valentini, C., Sivunen, A., and Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Comput. Educ.* 188:104564. doi: 10.1016/j.compedu.2022.104564

Sondermann, C., and Merkt, M. (2023). What is the effect of talking heads in educational videos with different types of narrated slides? *Contemp. Educ. Psychol.* 74:102207. doi: 10.1016/j.cedpsych.2023.102207

Sweller, J. (2011). "Cognitive load theory," in *The psychology of learning and motivation: Cognition in education*, Vol. 55, eds J. P. Mestre and B. H. Ross (Cambridge, MA: Elsevier Academic Press), 37–76. doi: 10.1016/B978-0-12-387691-1.00002-8

Sweller, J., Van Merrienboer, J. J. G., and Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205

Tao, Y., Zhang, G., Zhang, D., Wang, F., Zhou, Y., and Xu, T. (2022). Exploring persona characteristics in learning: A review study of pedagogical agents. *Procedia Comput. Sci.* 201, 87–94. doi: 10.1016/j.procs.2022.03.014

Veletsianos, G. (2010). Contextually relevant pedagogical agents: Visual appearance, stereotypes, and first impressions and their impact on learning. *Comput. Educ.* 55, 576–585. doi: 10.1016/j.compedu.2010.02.019

Wang, Y., Gong, S., Cao, Y., and Fan, W. (2023). The power of affective pedagogical agent and self-explanation in computer-based learning. *Comput. Educ.* 195:104723. doi: 10.1016/j.compedu.2022.104723

Appendix B

Manuscript 2

Decker, D., & Merkt, M. (2024). Shake it or light it! The effects of cueing in desktop-VR learning environments on search time and learning. *Journal of Computer Assisted Learning*, 40(3), 1201–1217. https://doi.org/https://doi.org/10.1111/jcal.12945

Author	Author Position	Scientific ideas %	Data generation %	Analysis and interpretation %	Paper writing %
Daniela Decker	1	75	100	90	80
Martin Merkt	2	25	0	10	20

Received: 20 July 2023 | Revised: 17 November 2023 | Accepted: 14 January 2024

DOI: 10.1111/jcal.12945

ARTICLE

Journal of Computer Assisted Learning WILEY

Shake it or light it! The effects of cueing in desktop-VR learning environments on search time and learning

Daniela Decker | Martin Merkt

Department Teaching, Learning, Counselling, German Institute for Adult Education – Leibniz Centre for Lifelong Learning, Bonn, Germany

Correspondence
Daniela Decker, Department Teaching, Learning, Counselling, German Institute for Adult Education, Leibniz Centre for Lifelong Learning, Heinemannstraße 12-14, 53175 Bonn, Germany.
Email: decker@die-bonn.de

Abstract

Background: Virtual reality (VR) offers much potential for learning, but it challenges learners' orientation.

Objectives: This paper investigates whether it is possible to use light or movement cues to facilitate orientation in a search task in a desktop-VR environment so that participants can better attend to the learning content presented simultaneously.

Methods: In two pre-registered online experiments, we investigated the effects of cueing (light and movement) on search time, learning, and several evaluation variables. Participants were asked to find tools in a virtual workshop, while information about the respective tool was narrated. Experiment 1 ($N = 60$) used a within-subject design, that is, the objects were alternately highlighted by light, movement or not. For Experiment 2 ($N = 159$) the narration was substantially shortened, and a between-subject design was used. Cognitive load and presence were measured additionally.

Results and conclusions: In Experiment 1, only the movement cue decreased search time, indicating automatic guidance of learners' attention. There was no effect of cueing on learning, which may be due to the average search time being substantially shorter than the narration, leaving sufficient time to attend to the narration exclusively. In Experiment 2 search times were significantly faster for both cueing methods, but only the light cue resulted in better learning outcomes, which could be explained by the slightly lower presence in the movement cue condition.

Implications: Results imply that it is important to develop cues that automatically guide attention without reducing presence.

KEYWORDS
cueing, dual-task paradigm, presence, signalling, virtual reality learning

1 | INTRODUCTION

Since Virtual Reality (VR) Systems have become more affordable, using them in an educational context has become more feasible (Parong & Mayer, 2018; Radianti et al., 2020). However, to make the most of the possibilities of VR in educational settings, the learning materials should be designed to overcome initial challenges, such as user disorientation, which may contribute to suboptimal learning outcomes with VR. Whereas VR is generally defined as a computer-generated environment (Ardiny & Khanmirza, 2018), which can either

be displayed on a computer screen (desktop-VR), or in a head-mounted display (Buttussi & Chittaro, 2018), this manuscript focuses on desktop-VR.

1.1 | Theoretical background

VR learning environments are often described as discovery-based and focus on learning through active participation. Being immersed in a graphically rich virtual environment offers great potential to facilitate learning, can help to create a connection between existing and new knowledge and may encourage learners to select, organise and integrate information themselves (Araiza-Alba et al., 2021). Further, interactive learning environments offer direct feedback for learners' actions (Makransky et al., 2020; Mikropoulos & Natsis, 2011). Implementing educational scenarios in VR environments offers the possibility to simulate and repeatedly practice scenarios, which would be expensive or dangerous in real life. This can reduce risks and save costs (Arnaldi et al., 2018). Additionally, several studies imply that intrinsic motivation (Cheng & Tsai, 2019; Makransky & Lilleholt, 2018; Olmos-Raya et al., 2018) and enjoyment (Meyer et al., 2019) were higher in VR learning environments.

Despite all these advantages, the use of VR for learning is not necessarily beneficial for the learning results, as it may overload or distract students (Makransky et al., 2019). For example, cognitive load theory (CLT, Sweller et al., 1998) postulates that cognitive resources in working memory are limited. Thus, learning materials should present information in a way that avoids cognitive overload. CLT distinguishes between three types of cognitive load, namely intrinsic, extraneous, and germane cognitive load (Sweller et al., 1998). Intrinsic cognitive load is inherent to the complexity of the task and depends on the number of elements the learner must deal with simultaneously and the learners' prior knowledge. Germane cognitive load represents deliberately invested mental effort to actively deal with the learning materials in order to construct a comprehensive mental representation. In contrast, the non-optimal design of the learning materials and instructions results in extraneous cognitive load (Sweller et al., 1998). Whenever learners must invest cognitive resources that are irrelevant to the actual task due to subpar instructional design, extraneous cognitive load is increased and may contribute to cognitive overload (Klepsch et al., 2017).

The Cognitive Affective Model of Immersive Learning (CAMIL, Makransky & Petersen, 2021) was developed to deal with the cognitive requirements of learning in VR, especially concerning instructional design and methods. Most importantly, CAMIL assumes that immersion, control factors, and representational fidelity positively affect presence and that control factors moreover positively affect agency (Makransky & Petersen, 2021). Immersion is a measure of the liveness of a system and the extent to which the system can exclude the external environment (Cummings & Bailenson, 2016), variables such as degree, immediacy, and type of control constitute the control factors (Witmer & Singer, 1998), and accuracy of the representation can be defined as the extent to which the representation corresponds to the real world (Dalgarno & Lee, 2010; Makransky & Petersen, 2021). CAMIL emphasises that presence and agency are general psychological prerequisites for learning in VR (Makransky & Petersen, 2021). Presence is defined as the psychological feeling of being there in a VR world, that is, being in a place or environment even if you are physically in a different place (Lee, 2004; Witmer & Singer, 1998). Agency is defined as the sense of creating and controlling actions (Moore & Fletcher, 2012), the most important predictor of which is that users can exercise control over their actions in the environment (Johnson-Glenberg, 2019).

Next to these positive effects of immersion, research also suggests that more immersion often leads to higher levels of extraneous cognitive load (Makransky et al., 2019; Meyer et al., 2019; Parong & Mayer, 2021). Therefore, to make optimal use of immersive VR learning environments, learning materials should be designed to avoid an additional increase in extraneous cognitive load (Makransky & Petersen, 2021). In particular, CAMIL emphasises that searching for relevant content might increase the extraneous cognitive load (Makransky & Petersen, 2021). In a VR environment, learners are surrounded by the virtual world and must decide in which direction they look. Consequently, they only perceive content in their limited field of view (FOV, Zhu et al., 2020). The FOV varies depending on the learners' position in the environment and can be understood as the maximum angle, which is currently visible (Shi et al., 2017). Thus, areas of interest could be located outside the current field of view, which makes the guidance of attention more important, but also more challenging (Grogorick et al., 2018; Makransky et al., 2021). A lack of guiding cues might therefore increase extraneous cognitive load (Makransky & Petersen, 2021), which in turn might lead to user disorientation and difficulties in navigation, which were observed in several studies (Marsh & Smith, 2001; Smith & Marsh, 2004). Therefore, this study focuses on how to improve the guidance of attention to facilitate orientation with two different cueing methods.

1.2 | Cueing in VR learning environments

In the following section, we will discuss cueing as a method that could potentially help overcome user disorientation. Cueing is defined as an approach to solving problems of information extraction by guiding learners to the currently most important information (Boucheix et al., 2013; van Gog, 2021). The cueing (or signalling) principle states that the use of cues in learning materials can draw learners' attention to relevant information, highlight important content and thus facilitate learning (Alpizar et al., 2020; Richter et al., 2016; Schneider et al., 2018; van Gog, 2021). Cueing is used in various learning materials ranging from static to animated materials (Boucheix et al., 2013; de Koning et al., 2011; Lin & Atkinson, 2011). Highlighting key points in learning material is necessary since the core processing of the visual system is limited to information located in the foveal vision. Thus, only the parts on which the learner focuses receive sufficient attention for further cognitive processing (de Koning et al., 2010). Therefore, cueing is especially recommended for the first encounters with multimedia learning material (van Gog, 2021).

To promote information extraction in conventional learning videos, the range of usable options for attention guidance might be broader due to the predefined field of view (e.g., see Schneider et al., 2018). In VR learning environments, where learners are surrounded by the learning environment and have to determine the viewing direction themselves (Rothe & Hußmann, 2018), finding the direction with the most relevant content can be more challenging (Makransky et al., 2021). Thus, drawing attention to the important parts of a VR environment could be more difficult than in conventional materials. Nevertheless, VR designers have adopted various approaches from conventional information guidance methods (such as adding circles and arrows) and have begun to test them for VR films. These approaches have not yet been fully developed for interactive VR (Speicher et al., 2019) or uniformly defined. These are often referred to as cueing, signalling, (subtle) gaze direction or visual guidance (Seok et al., 2021) and differ in their effectiveness and obtrusiveness.

Rothe and Hußmann (2018) differentiate between non-diegetic and diegetic cues. Whereas non-diegetic cues add elements to the scene (e.g., arrows), diegetic cues are integrated into the actual scene. Non-diegetic methods such as taking the viewer directly to the target area via autopilot or providing arrows that direct the viewer's gaze have already been studied in VR videos and could both ease orientation (Lin et al., 2017). Further, non-diegetic methods like visual text cues and arrows had positive effects on learning outcomes compared to conditions without cues in VR learning videos (Liu et al., 2022) and VR environments (Albus et al., 2021). Albus and Seufert (2022) showed that in a DVR learning environment, cueing with green circles improved learning and decreased cognitive load. However, the implemented cues did not assist learners in freely navigating through the learning materials. In contrast, the cues directed learners' attention to elements (i.e., animals) that they were already attending to. In particular, as soon as participants clicked on an animal a narration started and cues highlighted relevant parts of these animals (e.g., ears, hoofs). However, adding non-diegetic cues as additional elements to the learning materials may facilitate information extraction, but since it might also reduce learners' presence, which was described as an important prerequisite for successful learning in VR (see CAMIL, Makransky & Petersen, 2021), we focus on diegetic cues in our study.

In contrast to non-diegetic cues, diegetic cues are integrated into the learning scene and may thus be better suited to maintain learners' presence in the learning materials. In this regard, Rothe and Hußmann (2018) tested the effectiveness of three different types of diegetic cues: light, movement, and sound with a within-subject design in which all participants watched the same VR movie. Using heat maps that display viewing direction combined with statistical analysis for spatial data, they found out that all cueing types of cues were effective to highlight relevant objects. Since sound cues could interfere with the spoken learning content in our learning environment, we decided to only include light and movement cues in the present study. However, Rothe and Hußmann (2018) demonstrated the positive effects of light and movement cues on user orientation but did not address the potential benefits of these two cueing methods on learning. From a theoretical perspective, becoming oriented in a virtual learning environment while at the same time attending to learning materials may be considered a manifestation of dual tasking which has been shown to challenge learners (Bender et al., 2017; Hornof et al., 2010; Meyer & Kieras, 1997; Pashler, 1994). The burden of dual-tasking may be relieved by facilitating performance in one task (Bender et al., 2017; Dux et al., 2009; Hazeltine et al., 2002; Strobach et al., 2012). In the same vein, facilitating orientation in VR by means of light and movement cues might free up cognitive resources that learners might invest in processing the learning materials. Thus, it is possible that if one of those two tasks (the search task) is supported by cueing, it might also result in better learning by reducing the conflict for resources. This question will be investigated in our study, where the learning content is narrated during the search task and therefore constitutes as a dual task.

1.2.1 | Potential negative effects of cueing support on the formation of a spatial representation

Since the spread of digital technology, researchers have debated the impact of these technologies on the cognitive abilities of their users. Negative effects of information technologies on memory capacity have become known as the 'Google effect' (Sparrow et al., 2011), which can be broadly defined as not encoding or as forgetting information that has been digitally stored (van der Haak, 2014). Whether this is intentional because people consciously rely on digital storage devices (van der Haak, 2014), or due to a lack of practice in encoding (Spitzer, 2014), is a matter of debate. However, recent research has shown that technological support that facilitates performance in a practice phase may hamper performance in a test phase (Moritz et al., 2020). In Experiment 1, Moritz et al. (2020) had participants extract information from graphical data representations such as maps. Whereas half of the participants were allowed to manipulate the graphical representations so that they could extract information easily, the other half was not allowed to do so during the practice phase. None of the participants was allowed to manipulate the graphical representations during the test phase. Whereas participants who could manipulate the graphical representations performed better in the practice phase than participants who were not allowed to manipulate the representation, they dropped to the performance of an untrained baseline control group in the delayed test phase, implying that they were not able to maintain their superior performance levels. Similarly, several studies suggest that navigational aid is detrimental to the development of spatial representations, but the underlying effects are still in discussion (Dahmani & Bohbot, 2020; Gardony et al., 2015; Hejtmánek et al., 2018; Ishikawa et al., 2008). For example, Ishikawa et al. (2008) compared the orientation and spatial representation of participants in three groups who navigated a route using either GPS, a map, or their own experience (i.e., they were first guided by the study leader). The participants had to walk from a starting point to a destination in an unfamiliar city. They found that the participants who used GPS walked significantly longer, and their maps of their routes were

less accurate than those of the participants who had to navigate without GPS.

Therefore, we want to investigate whether the support provided by cues in a VR learning environment also has detrimental effects on the formation of a spatial mental representation. Specifically, learners might rely on cueing when searching for relevant information and thus develop an inadequate mental spatial representation of the environment.

1.3 | Overview of the experiments

To address the open questions outlined above, we conducted two experiments that investigate whether cueing with light and movement reduces search time and supports learning compared to a condition without cueing. Further, Experiment 1 (within-subject design) assessed learners' mental spatial representations of the item's location in the learning environment, whereas Experiment 2 (between-subject design) assessed learners' perceptions of presence and cognitive load.

2 | EXPERIMENT 1

2.1 | Hypotheses

In line with the signalling principle (Alpizar et al., 2020; Schneider et al., 2018; van Gog, 2021), multiple studies have demonstrated that cueing positively affects learning outcomes in animated learning videos with a predefined field of view (de Koning et al., 2007, 2011; Lin & Atkinson, 2011). However, there is limited research on the effects of cueing in VR learning environments where learners choose their own field of view for solving a search task with simultaneous learning content. Addressing these open research questions, we pre-registered the following hypotheses on OSF.[1]

Since Rothe and Hußmann (2018) found that light and movement cues improved search efficiency in VR, we hypothesise that light and movement cues will lead to faster search times in a desktop-VR learning environment. Moreover, we assume that the movement cue is more effective than the light cue (Hypothesis 1).

Moreover, Albus and Seufert (2022) showed positive effects of cueing on learning in desktop-VR learning environments where green circles highlighted the part of a virtual animal in a forest while learning content about that part was narrated. Since the narrative in this study only began when the animal was already in the learners' focus, the cueing assisted the learners in associating the narrated information with the corresponding part of the animal, but the purpose of these cues was not to support initial orientation in the 360° environment. However, in our study, the narration starts at the beginning of the task and the learner must find the corresponding 3D object simultaneously, constituting a dual task. Thus, neither the cueing method used by Albus and Seufert (2022), nor the one developed by Rothe and Hußmann (2018), which highlighted objects using movement and light, have been tested in a dual-task learning scenario. Thus, since research on dual-tasking suggests that facilitating performance in one task might facilitate overall performance (Hornof et al., 2010), we assume that learning outcomes will be better when the search task in the learning environment is supported by cueing (Hypothesis 2) because cues might free up learners' cognitive resources. We will further investigate whether there are differences in the effectiveness of the light and the movement cue exploratively.

Further, several studies showed that the use of navigational aids might impair spatial memory, which manifested in poorer performance in both navigation and landmark recall tasks (Gardony et al., 2015). However, since divided attention between navigation task and navigation aid was discussed to be one possible reason for these results, we investigated whether cueing with light and movement within the scene of the learning environment also impairs the formation of a long-term representation of the spatial layout. Adverse effects of cueing could be caused by learners relying on digital support, which could lead to less thorough encoding of spatial information (van der Haak, 2014). We therefore assume that the locations of the objects that had to be found without cueing are better remembered later (Hypothesis 3). Moreover, we will investigate whether there are differences between the cueing methods exploratively.

2.2 | Method

2.2.1 | Participants and design

The experiment was completed by 70 participants recruited using Prolific,[2] of which 10 had to be excluded due to the pre-defined exclusion criteria, namely non-termination of the learning environment (6 participants), technical problems during the learning environment (1 participant) and using search engines (3 participants) for the knowledge test. This resulted in a sample of 60 German-speaking participants (38 male, 22 female). Their mean age was 29.00 years ($SD = 7.62$). An a priori power analysis conducted with G*Power (Faul et al., 2007) with a fixed power of 0.95, a significance level of 0.05 and an expected effect size of 0.25 resulted in a minimum sample size of 43 for the intended analyses. The independent variable (within subjects) was cueing with the factor levels light, movement, and none. The participants were randomly assigned to one of the six sequences, which were used for balancing purposes and are not included as a factor in the analyses. Participants received 9.62€ for their participation.

2.2.2 | Learning materials

The desktop-VR learning environment represented a virtual wood workshop (see Figure 1), in that participants were asked to locate

[1] https://osf.io/vm6xj/?view_only=6c07aaea38d249fdaeb35638a1d068b4.

[2] https://prolific.co/.

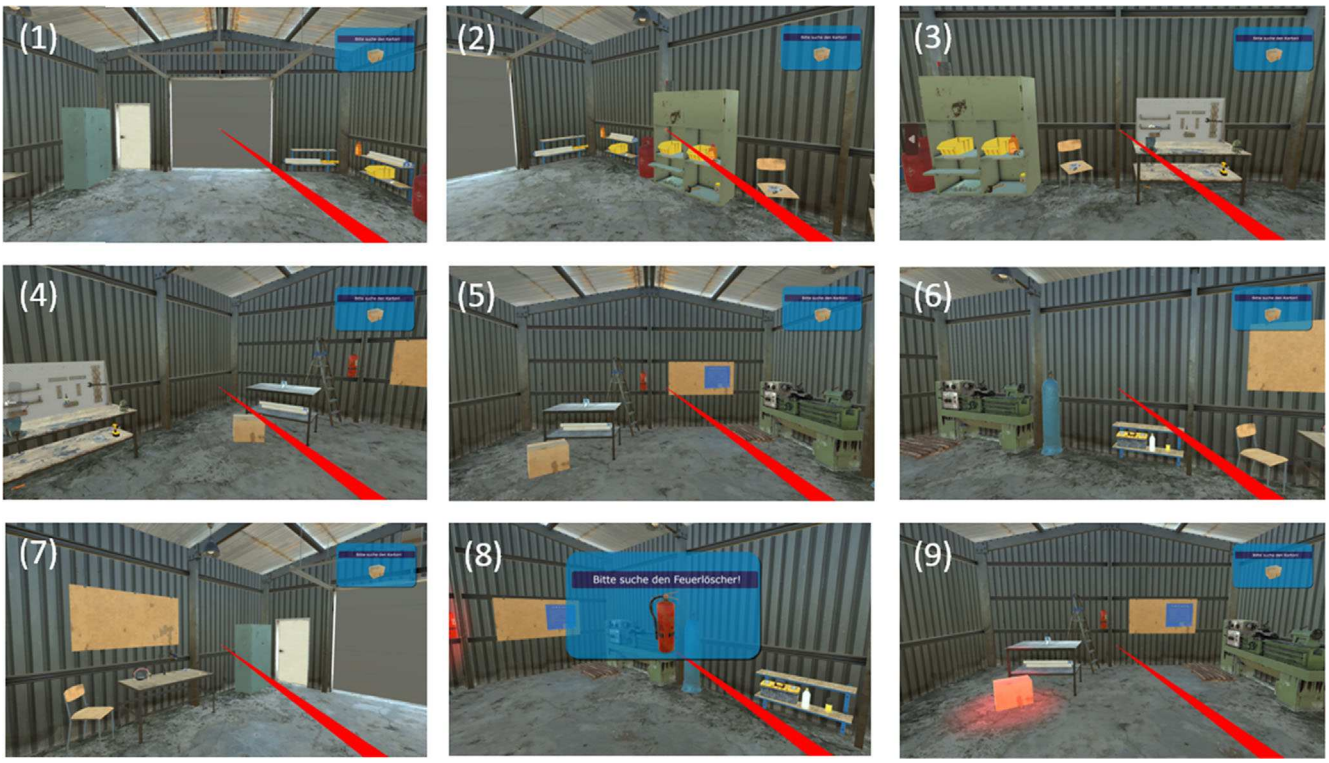


FIGURE 1 The learner's perspective of the learning environment. (1)–(7) Step-by-step sections of the learning environment in a clockwise direction; (8) Example of a UI with a search prompt; (9) Example of an object with light cue.

16 objects with a virtual laser pointer controlled by the PC mouse. The objects were distributed evenly throughout the learning environment. The first object was only for practice. Participants solved search tasks for various handicraft objects, such as brushes, pliers, and hammers, while a male human off-screen-narrator provided learning content about the respective objects (duration per object: between 27 and 75 s). Regarding the movement cue, the cued object moved minimally in three-dimensional space ("shook itself"). Regarding the light cue, the cued object was illuminated with a reddish light. Whereas each search task was presented only once per participant, we balanced the assignment of the search task to the cueing condition across participants, so that each object was represented with each cueing method across participants. To avoid sequence effects, these three resulting sequences were mirrored, resulting in a total of six sequences for the experiment. Search tasks were presented via a large user interface (UI) that showed a picture of the object in addition to a prompt, such as "Please find the hammer" for 4 s. After the large UI had disappeared, a smaller UI displaying the current task remained in the upper right corner. The selection of the correct object was confirmed by a UI and the learning environment was superimposed until the narration for the object was finished. If participants could not find an object after 300 s, the task was registered as missing value and the next task started, which did not occur in Experiment 1. Search times were automatically recorded and saved to a file on the users' desktop. The learning environment was created using Unity (Version 2019.4.13f1).

2.2.3 | Measures & instruments

Prior knowledge & interest

Participants' self-reported prior knowledge and interest in craft content were measured on a 7-point Likert scale ranging from 1 (almost no prior knowledge/interest at all) to 7 (a lot of prior knowledge/very great interest) with three items each. The topics were related to the contents of the learning materials and covered types and usage of different tools (saws, drills, hammers), weather protection of wood, and fire safety in workshops. Cronbach's alphas for the prior knowledge items and the interest items were 0.85 and 0.67, respectively. Further, participants' prior experience with VR (head mounted display and desktop-VR) and VR learning environments were queried using three items on the 7-point Likert scale ranging from 1 (almost no experience) to 7 (a lot of experience). Cronbach's alpha for the three VR experience items was 0.57.

Knowledge test

The learning outcome was recorded via 45 open items in a short answer format (three for each of the 15 search objects, excluding the practice trial). Participants had a maximum of 90 s for each question. Two independent raters assessed the correctness of the answers using a coding scheme, all Cohen's κ >0.79. One point was awarded per correctly answered question, resulting in a maximum score of 45 points. Partial points were awarded according to predefined criteria. Cronbach's alpha for the knowledge items was 0.83. The items

FIGURE 2 Mental spatial representation survey. Participants were asked which item was placed in the marked location (red square) during the search tasks.

included questions about factual information, such as "What is the mine of a pencil made of?" (correct answer: graphite).

Search time

During the search task, the learning environment recorded the search times (in seconds[3]) for each object (i.e., seconds elapsed between the start of the task until the selection of the correct object) and exported them into a result sheet created on the participants' computer desktop.

Mental spatial representation

To capture the mental spatial representation of the environment, participants were confronted with 15 screenshots (one per object) of the environment, in which the 15 search objects were not present. Participants were then asked which of the objects was in the marked location in the screenshot (see Figure 2).

2.2.4 | Procedure

The experiment was run online using LimeSurvey with participants recruited via Prolific. After giving informed consent, participants were randomly assigned to one of six groups based on the balanced sequences and started downloading the learning environment to their PCs. Afterwards a picture and sound test, prior knowledge and interest in the learning content as well as VR experience were assessed. Participants were then instructed on how to start and run the learning environment. After finishing the environment, participants were instructed to upload their result sheets and were asked to evaluate their learning experience.[4] Next, the knowledge test was administered, which contained 45 open items covering the narrated learning content and afterwards, participants' mental spatial representation of the spatial structure of the VR environment was measured. Finally, demographic data was queried, and the debriefing took place. In the debriefing, participants were informed about the research questions and the study design (see Figure 3).

2.3 | Results

After recording the data from the six different balanced sequences, we conducted ANOVAs (repeated measures) with cueing as a within-subject factor (light, movement, none) for each of the dependent variables. Descriptive data is presented in Table 1. The data and the analysis script are available on OSF.[5]

2.3.1 | Interest, prior knowledge, and VR experience

Since we used a within-subject design and did not assess prior knowledge and interest for every search object, we only reported descriptive data for the respective variables. Self-reported interest ($M = 3.05$; $SD = 1.04$), prior knowledge about the contents ($M = 2.37$; $SD = 1.19$), and self-reported VR

[3]This unit (i.e., seconds) was used because it ensures reliable measures using the Unity Delta Time on computers with different performance levels that display the learning environment with different FPS numbers.

[4]Since we used a within-subject design and did not assess the assessment variables after each trial, we cannot test for differences between different cueing conditions. Therefore, these variables mainly serve for intern evaluation purposes and will not be further reported.

[5]https://osf.io/2f87s/?view_only=173e32e228e84510b8f293a0dcd92898.

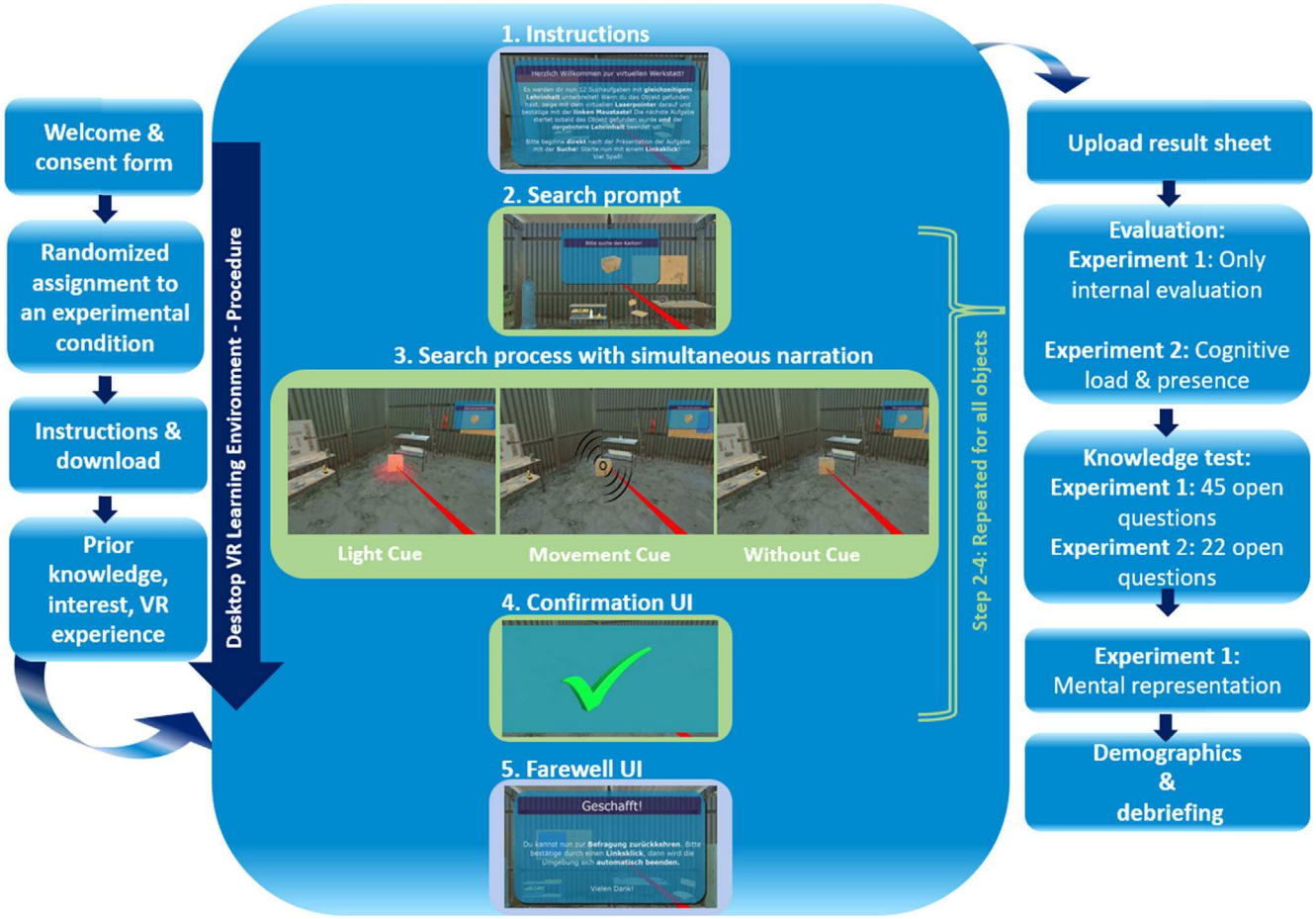


FIGURE 3 Procedure of the experiments. Experiment 1 repeated this procedure (steps 2–4) for 16 objects and manipulated cueing within subjects. Experiment 2 repeated this procedure for 12 objects and manipulated cueing between subjects.

TABLE 1 Experiment 1(within-subject design): Overall means and standard deviations.

	Cueing method			
Variable	**Light**	**Move**	**None**	**Overall**
Prior interest				3.05 (1.04)
Prior knowledge				2.37 (1.19)
VR experience				2.88 (1.16)
Search time	10.96 (6.98)	9.29 (5.30)	10.58 (5.94)	10.28 (5.67)
Knowledge test	6.34 (2.84)	6.58 (2.41)	6.47 (2.91)	19.39 (6.59)
Mental representation	2.45 (1.24)	2.46 (1.29)	2.57 (1.28)	7.48 (2.19)
N	20	20	20	60

experience ($M = 2.88$; $SD = 1.16$) were all on a low to medium level.

2.3.2 | Search time

Regarding Hypothesis 1, the ANOVA revealed a significant effect of cueing for search time, $F(2, 118) = 5.96$, $p = 0.005$, $\eta_p^2 = 0.09$. Bonferroni-adjusted post hoc comparisons revealed that search times were significantly faster when the objects were highlighted with the movement cue ($M = 9.29$; $SD = 5.30$) compared to the light cue ($M = 10.96$; $SD = 6.98$), $p = 0.010$, and the condition without cueing ($M = 10.58$; $SD = 5.94$), $p = 0.007$. There was no significant difference between the light cue condition and the condition without cues, $p = 1.000$. Therefore, the first hypothesis can be confirmed for the movement cue, but not for the light cue (see Figure 4).

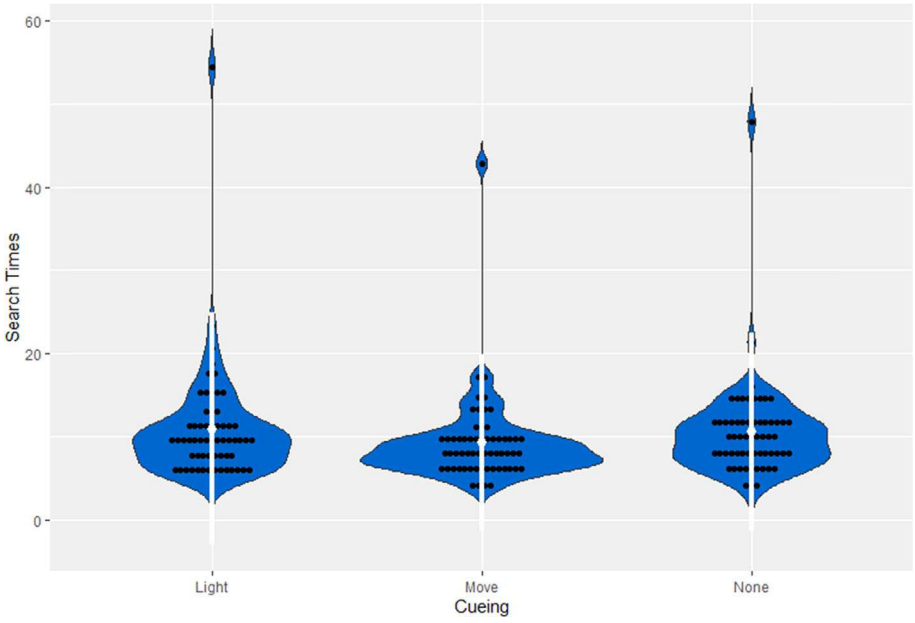


FIGURE 4 Experiment 1 – Search times. Violin plot for the search times in Experiment 1 (within-subject design) including the means and standard deviations represented by the white line.

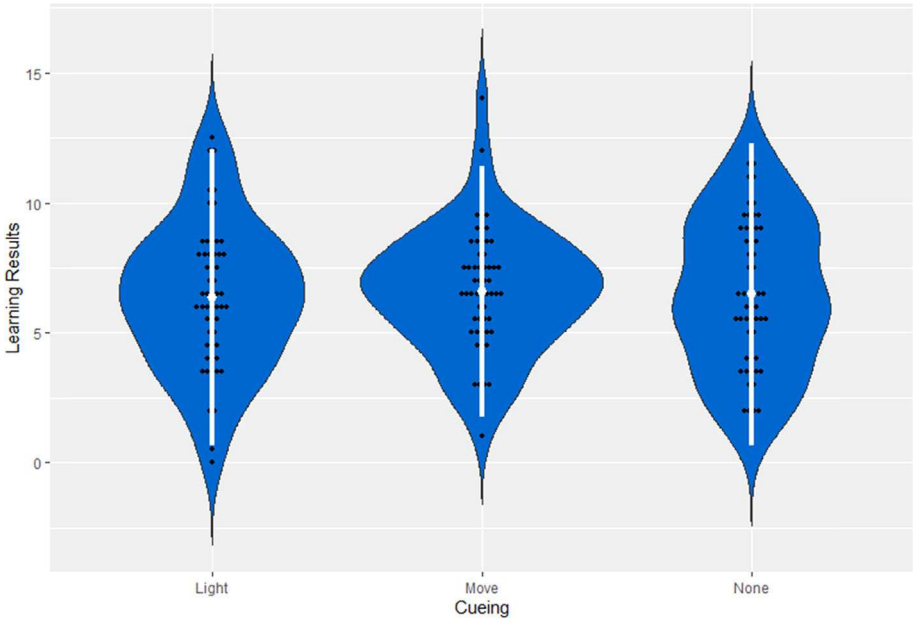


FIGURE 5 Experiment 1 – Learning Results. Violin plot for the learning results in Experiment 1 (within-subject design) including the means and standard deviations represented by the white line.

2.3.3 | Learning outcomes

Regarding Hypothesis 2, the ANOVA revealed no significant effect of cueing for the learning outcomes ($M = 19.39$, $SD = 6.59$), $F < 1$ (see Figure 5).

2.3.4 | Mental spatial representation

Concerning Hypothesis 3, the ANOVA revealed no significant effect of cueing for the mental spatial representation score ($M = 7.48$, $SD = 2.19$), $F < 1$.

2.4 | Discussion

Experiment 1 investigated whether different types of cueing (i.e., light cues and movement cues) affect participants' search times, learning outcomes, and mental spatial representation of the layout in a desktop-VR learning environment. Partially corroborating Hypothesis 1, we observed shorter search times for movement cues compared to the other two conditions; however, light cues unexpectedly did not facilitate searching. Possibly, this finding may be due to the more static nature of the light cues that may have attracted less attention than the dynamic movement cues, which seem to attract attention

automatically (Jiang et al., 2018). Especially in a within-subject design in that participants are confronted with different cues from trial to trial, the lack of predictability may be considered a disadvantage for cues that do not attract attention automatically. Therefore, for Experiment 2 we changed the design to a between-subject design.

Regarding Hypothesis 2, there was no effect of cueing on learning outcomes, which could be explained by the search times. Since all objects were found rather quickly in all three conditions ($M_{\text{Search Time}} = 10.27$, $SD = 5.67$), all participants had sufficient time to engage exclusively with the remaining narrated content that took between 27 and 75 s ($M = 44.13$; $SD = 13.30$). Accordingly, the overlap between the search task and narration, which would constitute a dual task, was not as large as we had initially expected. It is possible that cueing might have a positive effect on learning when the two tasks must be solved in parallel for a longer time and thus compete harder for the same cognitive resources. Thus, the narration for the search objects was shortened for Experiment 2 to increase its overlap with the search task.

Regarding Hypothesis 3, the lack of an effect of cueing on the mental spatial representation of the learning environment could be explained by the within-subject design that was administered in Experiment 1. In particular, all participants were confronted with five search tasks of each type (plus the practice trial), including five trials in which participants had to search for objects without cueing support. Thus, it is possible that participants encoded the locations of other objects they encountered in the environment while searching for the relevant search objects in the trials without cueing support. However, it should be noted that the overall average mental spatial representation score was at a low level ($M = 7.48$; $SD = 2.19$). This result could be explained by the short average search times since the learning environment was blended after the correct object was selected. In addition, assessing learners' mental spatial representations of a dynamic learning environment using static screenshots may have been suboptimal due to a mismatch between the learning and the testing context. To some extent, learners may have relied on the dynamic information when encoding the object locations. Accordingly, future studies could measure learners' mental spatial representation with a dynamic test.

In summary, the dynamic movement cue efficiently directed learners' attention, but the less salient light cue did not. However, none of the cueing methods affected learning outcomes, possibly due to a relatively small overlap of the search tasks and the narrated learning materials. To address these potential explanations, we implemented a between-subject design in Experiment 2 to test whether the light cue would also facilitate search times when the cues are predictable. Additionally, the narrated learning materials for Experiment 2 were significantly shortened, resulting in a stronger emphasis on factual knowledge. Finally, to get some insights into mechanisms underlying potential cueing effects, we investigated how cognitive load and presence were influenced by the two cueing methods.

3 | EXPERIMENT 2

3.1 | Hypotheses

Based on the potential explanations for the findings in Experiment 1, we adapted the design and learning material for Experiment 2 and preregistered the following hypotheses on OSF.[6]

> **Hypothesis 1.** The search task is solved faster if objects are highlighted by the movement cue.
>
> **Hypothesis 2.** Learning outcomes will be better when the search task in the learning environment is supported by the movement cue.
>
> **Hypothesis 3.** Extraneous Cognitive Load (environmental scale, see Andersen & Makransky, 2021) will be lower in the conditions with the movement cue.

Because the light cue did not affect search time in Experiment 1, we explored whether the light cue is more effective in a between-subject design concerning search time and learning outcomes. Also, the effects of cueing on presence, intrinsic cognitive load, and germane cognitive load are analysed exploratively.

3.2 | Method

3.2.1 | Participants & design

The experiment was conducted with 179 participants recruited via Prolific. Twenty participants had to be excluded due to the pre-defined exclusion criteria, namely using search engines for answering the knowledge test (10 participants), taking notes during the learning environment (6 participants), technical problems with the learning environment (4 participants) or missing values in the search times (2 participants).[7] This resulted in a final sample of 159 German-speaking participants (80 male, 77 female, 2 diverse), which was the adequate sample size based on a priori power analysis conducted with G*Power (Faul et al., 2007) (expected effect size = 0.25, alpha, level = 0.05, Power = 0.80). Their mean age was 30.23 years ($SD = 9.11$). In this experiment, cueing was manipulated between subjects with the factor levels light, movement, and none. Participants were randomly assigned to one of the three conditions and received 6.24€ for their participation.

3.2.2 | Learning materials

The desktop-VR learning environment was similar to the one used in Experiment 1, with the following changes: Participants were asked to

[6] https://osf.io/9v4rk/?view_only=c8a4e050839542f1ba3ecb10c32e1176.

[7] Since participants could be excluded because of more than one criterion, the sum of the individual exclusion reasons differs from the total number.

only find 12 objects (11 search objects and 1 practice trial at the beginning) and the narrated learning material presented during the search tasks was shortened (8 to 16 s; $M = 12.91$; $SD = 2.34$) to increase the overlap between the search task and the learning content. Search times were recorded comparable to Experiment 1.

3.2.3 | Measures & instruments

Prior knowledge, interest, and VR experience

Prior knowledge, interest, and VR experience were assessed as in Experiment 1 (see Section 2.3.1). Cronbach's alphas were 0.77 for prior knowledge, 0.86 for interest, and 0.69 for VR experience.

Cognitive load

Cognitive load was assessed with the VR-specific scale by Andersen and Makransky (2021), which we adapted for our learning environment.[8] It measures the cognitive load on an 11-point Likert scale from 0 (does not apply at all) to 10 (fully applies). In addition to germane (4 items, e.g., "The virtual environment really enhanced my understanding of the topics covered.", Cronbach's alpha = 0.92) and intrinsic cognitive load (3 items, e.g., "The topic covered in the virtual environment was very complex.", Cronbach's alpha = 0.95), this scale distinguishes between three kinds of extraneous cognitive load, namely environmental extraneous cognitive load (4 items, e.g., "The elements in the virtual environment made the learning very unclear.", Cronbach's alpha = 0.76), extraneous cognitive load for instruction (3 items, e.g., "The explanations used in the virtual environment were very unclear.", Cronbach's alpha = 0.77), and extraneous cognitive load for interaction (4 items, e.g., "The interaction technique used in the virtual environment was very unclear.", Cronbach's alpha = 0.70).

Presence

We assessed physical (Cronbach's alpha = 0.84) and self-presence (Cronbach's alpha = 0.90) using items from the German version of the presence scale developed by Makransky et al. (2017). The German version (Volkmann et al., 2018) demonstrated high correlations with the Igroup Presence Questionnaire (IPQ, Schubert, 2003) in a subsequent study by Volkmann et al. (2020). Both dimensions (physical and self-presence) were assessed with 5 items each, which had to be rated on a 5-point Likert scale from 1 (does not apply at all) to 5 (fully applies).

Knowledge test

The learning outcome was assessed with 22 open items (two for each of the 11 search objects, excluding the practice trial). These questions were selected from the pool of questions from Experiment 1, excluding questions that queried information that were cut from the narration, or which were too easy or difficult. Two independent raters used a coding scheme to assess whether the answers were correct (Cohens κ >0.80) and were awarded one point per correctly answered question, resulting in a maximum score of 22 points. Partial points were awarded according to predefined criteria. Cronbach's alpha for the knowledge items was 0.70.

3.2.4 | Procedure

This experiment was run online using LimeSurvey with participants recruited via Prolific. After giving informed consent, participants were randomly assigned to one of three groups based on the three cueing conditions (light, movement, none) and started downloading the learning environment to their PC. After a picture and sound test, prior knowledge and interest in the learning content as well as VR experience were assessed. Participants were then instructed on how to start and run the learning environment. After finishing the environment, participants were instructed to upload their result sheet, before presence and cognitive load were measured. Next, the knowledge test was administered. Finally, demographic data was queried, and the debriefing took place. In the debriefing, participants were informed about the research questions and the study design.

3.3 | Results

Data were analysed with ANOVAs with cueing as a between-subject factor (light, movement, none) for each of the dependent variables. The significance level was set at $\alpha = 0.050$. Descriptive data is presented in Table 2. The data and the analysis script are available on OSF.[9]

3.3.1 | Interest, prior knowledge, and VR experience

There were no differences between the three groups concerning learners' self-reported interest ($M = 2.95$; $SD = 1.39$), $F(2, 156) = 2.40$, $p = 0.094$, $\eta_p^2 = 0.03$; self-reported prior knowledge about the contents ($M = 2.42$; $SD = 1.19$), $F(2, 156) = 1.04$, $p = 0.352$, $\eta_p^2 = 0.01$; or self-reported VR experience ($M = 3.13$; $SD = 1.37$), $F < 1$. Thus, we can assume that these groups were comparable before the experiment.

3.3.2 | Search time

In line with Hypothesis 1, the ANOVA revealed a significant main effect of cueing for the average search time, $F(2, 156) = 13.03$, $p < 0.001$, $\eta_p^2 = 0.14$. Bonferroni-adjusted post-hoc tests showed faster search times for the light cueing condition ($M = 9.16$; $SD = 4.25$)

[8]The original instrument covered laboratory safety accordingly, we adapted it the to the virtual workshop in our study. Also, instead of the term simulation, we used the term virtual environment, which was used in the extraneous cognitive load for environment scale, as this term was better suited for our environment.

[9]https://osf.io/2f87s/?view_only=173e32e228e84510b8f293a0dcd92898.

TABLE 2 Experiment 2 (between-subject design): Overall means and standard deviations.

Variable	Cueing method			Overall
	Light	Move	None	
Prior interest	3.02 (1.40)	2.63 (1.32)	3.20 (1.41)	2.95 (1.39)
Prior knowledge	2.39 (1.26)	2.28 (1.12)	2.61 (1.17)	2.43 (1.19)
VR experience	2.94 (1.34)	3.17 (1.32)	3.30 (1.44)	3.14 (1.37)
Search time	9.16 (4.25)	8.99 (5.42)	13.09 (4.27)	10.41 (5.02)
Knowledge test	9.78 (3.90)	8.15 (3.08)	8.35 (3.30)	8.76 (3.50)
Physical presence	2.56 (0.78)	2.34 (0.84)	2.73 (0.75)	2.54 (0.80)
Self-presence	1.75 (0.79)	1.67 (0.75)	1.88 (0.82)	1.77 (0.79)
Intrinsic cognitive load	4.19 (2.37)	3.87 (2.44)	4.04 (2.08)	4.04 (2.29)
Environmental extraneous cognitive load	3.83 (2.03)	3.49 (1.97)	3.33 (2.05)	3.55 (2.01)
Extraneous cognitive load for instruction	3.23 (1.90)	3.35 (2.01)	3.23 (2.02)	3.27 (1.96)
Extraneous cognitive load for interaction	3.50 (2.00)	3.39 (1.98)	2.92 (1.90)	3.27 (1.96)
Germane cognitive load	2.83 (2.29)	2.79 (2.09)	3.33 (2.01)	2.98 (2.13)
N	53	53	53	159

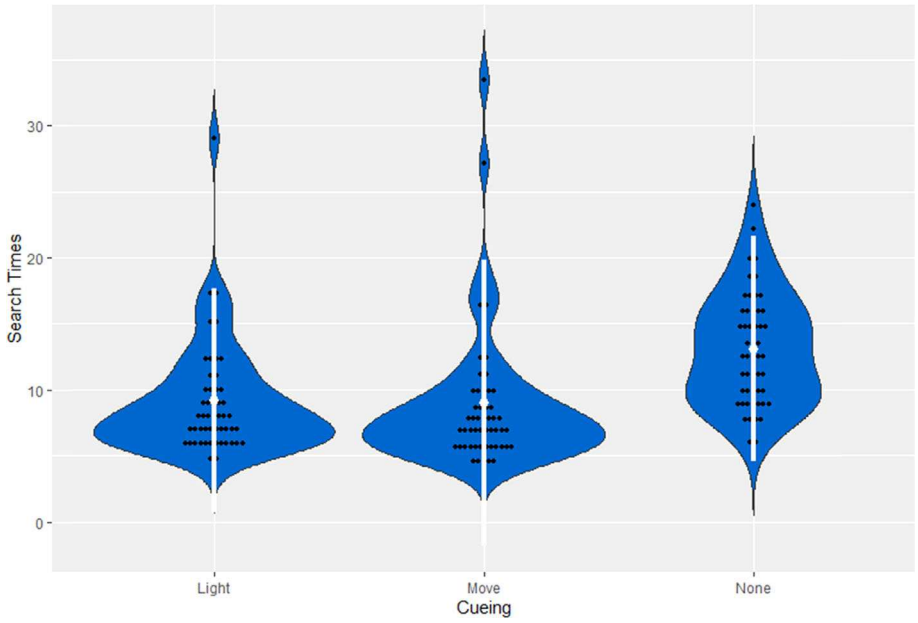


FIGURE 6 Experiment 2 – Search times. *Source*: Violin plot for the search times in Experiment 2 (between-subject design) including the means and standard deviations represented by the white line.

than the condition without cueing ($M = 13.09$; $SD = 4.72$), as well as for the movement cueing condition ($M = 8.99$; $SD = 5.42$) compared to the condition without cueing, both $p < 0.001$. The light and movement conditions did not differ significantly, $p = 1.000$ (see Figure 6).

3.3.3 | Learning outcomes

Concerning Hypothesis 2, the ANOVA revealed a significant effect of cueing for the learning outcome, $F(2, 156) = 3.54$, $p = 0.031$, $\eta_p^2 = 0.04$. Bonferroni adjusted post-hoc tests revealed better learning outcomes for the condition with light cues ($M = 9.78$; $SD = 3.90$) compared to the condition with movement cues ($M = 8.15$; $SD = 3.08$), $p = 0.048$, whereas the light condition did not significantly differ from the condition without cueing ($M = 8.35$; $SD = 3.30$), $p = 0.10$. The movement cue condition and the condition without cueing also did not differ significantly, $p = 1.000$ (see Figure 7).

3.3.4 | Cognitive load

There was no significant effect of cueing for extraneous cognitive load for the environmental scale ($M = 3.55$, $SD = 2.01$), $F < 1$. Further exploratory analyses revealed no significant differences for the intrinsic cognitive load ($M = 4.04$, $SD = 2.29$), $F < 1$, extraneous cognitive load for instruction ($M = 3.27$, $SD = 1.96$), $F < 1$, extraneous cognitive load for interaction ($M = 3.27$, $SD = 1.96$), $F(2, 156) = 1.30$,

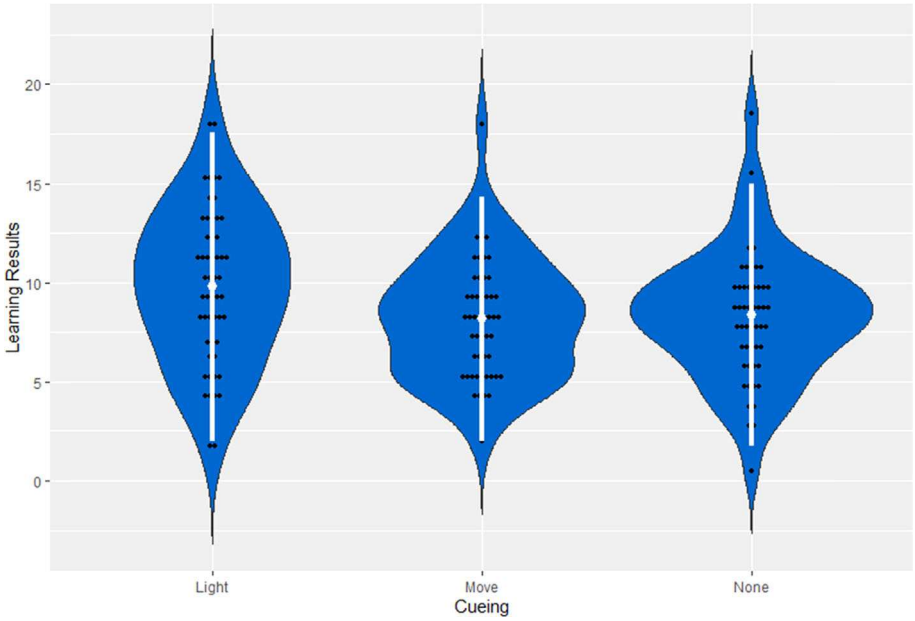


FIGURE 7 Experiment 2 – Learning Results. *Source*: Violin plot for the learning results in Experiment 2 (between-subject design) including the means and standard deviations represented by the white line.

$p = 0.275$, $\eta_p^2 = 0.02$, and germane cognitive load ($M = 2.98$, $SD = 2.13$), $F(2, 156) = 1.07$, $p = 0.345$, $\eta_p^2 = 0.01$.

3.3.5 | Presence

An exploratory analysis for presence revealed a significant effect of cueing on physical presence, $F(2, 156) = 3.29$, $p = 0.040$, $\eta_p^2 = 0.04$. Bonferroni-adjusted post-hoc tests revealed a significant difference between the movement condition ($M = 2.24$, $SD = 0.84$) and the condition without cues ($M = 2.73$, $SD = 0.75$), $p = 0.035$, but no significant difference to the light cue condition, $p = 0.427$. The scores for the light condition ($M = 2.56$, $SD = 0.78$) did not differ significantly from the condition without cues, $p = 0.843$. There was no main effect on self-presence, $F(2, 156) = 1.01$, $p = 0.366$, $\eta_p^2 = 0.01$.

3.4 | Discussion

Experiment 2 investigated whether different types of cueing (i.e., light cues and movement cues) affect participants' search times, learning outcomes, cognitive load, and presence in a desktop-VR learning environment. In line with Hypothesis 1 and replicating Experiment 1, the movement cue significantly reduced search time. Further, in this experiment, the light cue was also able to reduce search time. Thus, both cueing methods guided attention successfully.

Further, even though both types of cueing reduced search times, only the light cue resulted in better learning outcomes. The effects of different types of cueing on learning could be explained by differences in physical presence. The movement cue lowered perceived physical presence compared to the control condition without cueing, whereas the light cue did not negatively affect physical presence compared to the control condition. This observation may explain why the shorter search times translated to superior learning outcomes for the light cue, but not for the movement cue. However, it should also be mentioned that physical presence did not significantly differ between the two cueing conditions. Nevertheless, since CAMIL (Makransky & Petersen, 2021) states that presence is a prerequisite for learning in VR, the reduced presence in the condition with the movement cue might have hindered further learning success. A factor that influences presence is the accuracy of the representation (Dalgarno & Lee, 2010; Makransky & Petersen, 2021), that is, to what extent the representation corresponds to the real world. Since objects in the real world do not move automatically, the movement cue could have been disruptive and therefore created the perception of lower physical presence in our learning environment. However, illuminating objects with artificial light is quite common in the real world, which may have led to the light cue being perceived as more natural.

In contrast to Hypothesis 3, both the movement cue and the light cue did not significantly reduce environmental extraneous cognitive load. Overall, the level of perceived cognitive load was rather low, which could be explained by the design of the tasks. Although it was clearly specified that there would be a knowledge test about the narrated content, the prompts in the VR environment exclusively referred to the search tasks. This framing might have led to the search tasks being perceived as primary tasks, a problem that is well known in research on the dual-task paradigm and is criticised in many designs (Hazeltine et al., 2002). In addition, only two participants had to be excluded due to missing values in the search times, which in turn means that the remaining participants found all search objects relatively quickly, according to the mean search times over all objects ($M = 10.41$, $SD = 5.02$). Therefore, the participants might have perceived the search tasks as easy and therefore rated their intrinsic and extraneous cognitive load as rather low, as shown by the rather low scores in those subscales of the cognitive load scale. However, since Andersen and Makransky (2021) also found no consistent association

between cognitive load and retention, the differences in our learning results are more plausibly explained by the effects of presence. Other directions for future research will be discussed in section 4.2.

4 | GENERAL DISCUSSION

4.1 | Interpretation of the results

In two experiments, we investigated the effects of light and movement cues on search times, learning outcomes, mental spatial representation (Experiment 1), cognitive load, and presence (Experiment 2) in a desktop-VR learning environment. In both experiments, the movement cue successfully guided learners' attention and facilitated searching in the within and the between-subject design, whereas the light cue was only efficient in a between-subject setting, which afforded higher predictability for learners regarding which type of cue to expect. Because movement naturally attracts attention (Jiang et al., 2018), it is plausible to assume that the movement cues may guide attention in a more automatic way because they are more salient than static light cues.

Overall, the light cue positively affected both search times and learning outcomes in Experiment 2, but not in Experiment 1. This difference could be explained by switching from a within-subject design in Experiment 1 to a between-subject design in Experiment 2. Through this change, participants in Experiment 2 could better predict which type of cue to attend to. Consequently, because of the resulting facilitation of search times, the light cue may have freed cognitive resources so that participants' learning outcomes were improved by the light cue. In contrast, across both experiments, the movement cue facilitated search times, but it did not improve learning, which may be due to decreased physical presence compared to the control condition without cueing. The light cue, however, did not negatively affect physical presence compared to the control condition without cueing. This observation may explain why the shorter search times translated to superior learning outcomes for the light cue compared to the movement cue in Experiment 2.

Further, it should be noted that we shortened the narration of the learning material to increase the overlap between the search task and the learning materials, which might have increased task difficulty in Experiment 2. However, the presumably increased task difficulty did not automatically lead to better learning outcomes if the search times were shortened by cueing, but only if the presence was not reduced. To gain further insights into the effects of task difficulty, future research could systematically vary task difficulty and cueing in a single experiment.

Summing up, the costs and benefits of both types of cues investigated in these experiments need to be carefully weighed. The benefits of the movement cue shown in these experiments are the acceleration of search times in both predictable and unpredictable scenarios. It can thus be assumed that attention is directed automatically, which in turn should have reduced the extraneous cognitive load. However, this presumably lower cognitive load did not lead to better learning outcomes in either of our experiments, so it can be assumed that the reduction in presence, which is considered a prerequisite for successful learning, on the cost side apparently prevented better learning, based on the results of Experiment 2. Accordingly, from a usability perspective, movement cues would be more suitable when quick navigation is the task priority (e.g., in monitoring tasks that require immediate attention). The benefits of the light cue, however, are that it did not reduce presence and resulted in better learning results in a predictable between-subject design, but on the cost side, it needs to be emphasised that it has some limitations concerning the scenarios it can be used in. Since the light cues do not direct attention as automatically as the movement cue, it might be limited to scenarios in that learners can predict which type of cue to attend to in the next trial.

4.2 | Limitations and directions for future research

Exploring new options for cueing in VR might be worthwhile since both cueing methods come with advantages and disadvantages. Thus, future research should focus on designing cues that combine the strengths of the investigated cueing methods while avoiding potential caveats. Regarding potential future cueing methods, it may be a promising next step to include human-like pedagogical agents with a cueing function in the learning materials. Human-like agents might automatically guide learners' attention (Davis, 2018; Kizilcec et al., 2015; Wang et al., 2018). Moreover, pedagogical agents that provide the narration for the learning materials might be less likely to reduce learners' physical presence and might additionally increase the social presence in the learning environment (Schmidt et al., 2019; Schneider et al., 2022).

Further, the light cues only resulted in faster search times when the type of cue was predictable in a between-subject design (Experiment 2). Whether a dynamic light cue (e.g., flickering or switching on and off) might attract attention as automatically as the movement cue without reducing presence might therefore be a worthwhile research question. Particularly since the set size and the amount of distractors (clutter) influence the level of difficulty (Botch et al., 2023), the effects of cueing should be strengthened with larger set sizes and more distractors. However, the cueing methods tested in these experiments may have limitations regarding those possible areas of application. In learning environments where several objects are close together, our light cue (which simulates a spotlight and therefore illuminates a slightly larger area around the object in question) may be less accurate as it may illuminate several small objects at once. Thus, increasing the difficulty of the search task by including more distractors in order to identify stronger effects of cueing on search times and extraneous cognitive load (which was rather low in Experiment 2), might put the light cue at an additional disadvantage compared to the movement cue. To solve this problem, other types of light cues could be developed, such as light cones that become narrower when participants get closer to the searched object. In addition, not only could search objects be illuminated with an external projector but illumination starting from the object itself (turning up the colour luminosity) could also be tested as a different type of light cue.

In addition, the distribution of search times in Experiment 2 (see Figure 6) provides some interesting insights, which could be investigated in further research. Descriptively, it is noteworthy that search times in Experiment 2 in the condition without cueing are more distributed than search times in the conditions with cueing. A possible explanation for this could be that learners need to rely on their spatial abilities when they are not supported by cueing. More specifically, it is possible that learners with high spatial abilities find their way around particularly quickly and learners with lower spatial abilities take longer. Conversely, this interpretation would also mean that the cueing methods tested in these experiments could compensate for these differences in spatial abilities. Since previous research (Huk, 2006) has focused on the relationship between spatial abilities and learning with 3D models (such as molecules, etc.), it would be interesting for future research to investigate how spatial abilities affect navigation in virtual space, and which type of cueing is appropriate for learners with low or high spatial ability. Methods for measuring spatial ability that could be used for this purpose are, for example, the mental rotation test (Vandenberg & Kuse, 1978) or the Spatial Orientation Test (Friedman et al., 2020; Hegarty & Waller, 2004).

Furthermore, although the cognitive load in Experiment 2 was rather low, the learning outcomes in all conditions left room for improvement. As discussed in Section 3.4, one possible explanation might be the design of the task, where the search task might have been perceived as the primary task. Therefore, when participants were asked to rate their cognitive load after the learning environment but before the knowledge test, they may have rated their cognitive load based solely on their perception of the search task. Future research could potentially address this with a more nuanced measure of cognitive load, where participants rate their cognitive load twice, with the first assessment specifically relating to the search task and the second assessment specifically relating to the learning task. In addition, the cognitive load could also be assessed directly after the learning task and after the knowledge test to determine possible changes in participants' perceptions due to their performance in the knowledge test. As implementing both of these suggestions would result in a total of four measurements, this should be carefully considered from an efficiency perspective. Alternatively, future studies could also seek opportunities to assess the cognitive load in a within-subject design, where changes in cueing support between tasks might be more apparent to learners.

Further research on mental spatial representation should deal with the question of whether cueing hampers learners' mental spatial representations of the learning environment. However, a recent study from Stefanucci et al. (2022) suggests that cues in virtual environments might improve spatial orientation if those cues are placed inside the virtual environment because they do not divide attention like GPS Tools (which usually run on separate devices), which are usually placed outside the current field of view. However, it is important to note that the cues used in this study provided additional information for spatial orientation (such as distances) and they did not test simultaneous learning. Thus, even though we did not find any evidence for the detrimental effects of cueing on mental representation in Experiment 1, further research is necessary to rule out this possibility. Possible explanations for the non-significant effects of Experiment 1 were discussed in detail in Section 2.4 and could largely be explained by the floor effects, which could have been caused by the static assessment of learners' mental representations. Comparable to studies concerned with spatial orientation (Dahmani & Bohbot, 2020), mental spatial representation could be measured by presenting learners with a similar task as in the learning phase, but without cueing support, thus using a dynamic assessment of the spatial representation.

4.3 | Conclusion

To fully exploit the potential of VR learning environments, difficulties like user disorientation need to be addressed. Both methods tested in this study were capable of directing learners' attention; however, the light cue was only efficient in a between-subject design in that the type of cue was more predictable. The movement cue, however, also guided learners' attention in a less predictable within-subject design but reduced physical presence and hampered learning in comparison to the light cue. Therefore, future research should emphasise the design of cues that automatically guide learners' attention without reducing physical presence.

AUTHOR CONTRIBUTIONS

Daniela Decker: Conceptualization; writing – original draft; methodology; formal analysis; visualization. **Martin Merkt:** Conceptualization; writing – review and editing; supervision.

ACKNOWLEDGEMENTS

There was no external funding for the experiment. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available on OSF at: https://osf.io/2f87s/?view_only=173e32e228e84510b8f293a0dcd92898.

ETHICS STATEMENT

The experiment was approved by the local ethics committee of the German Institute for Adult Education in Bonn (DIE-LEK 2022-002). All participants gave informed consent prior to the experiment.

ORCID

Daniela Decker https://orcid.org/0000-0002-0811-8906
Martin Merkt https://orcid.org/0000-0001-8817-8433

REFERENCES

Albus, P., & Seufert, T. (2022). Signaling in 360° desktop virtual reality influences learning outcome and cognitive load. *Frontiers in Education*, *7*, 1–10. https://doi.org/10.3389/feduc.2022.916105

Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education*, *166*, 104154. https://doi.org/10.1016/j.compedu.2021.104154

Alpizar, D., Adesope, O. O., & Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educational Technology Research and Development*, *68*(5), 2095–2119. https://doi.org/10.1007/s11423-020-09748-7

Andersen, M. S., & Makransky, G. (2021). The validation and further development of a multidimensional cognitive load scale for virtual environments. *Journal of Computer Assisted Learning*, *37*(1), 183–196. https://doi.org/10.1111/jcal.12478

Araiza-Alba, P., Keane, T., Matthews, B., Simpson, K., Strugnell, G., Chen, W. S., & Kaufman, J. (2021). The potential of 360-degree virtual reality videos to teach water-safety skills to children. *Computers & Education*, *163*, 104096. https://doi.org/10.1016/j.compedu.2020.104096

Ardiny, H., & Khanmirza, E. (2018). The role of AR and VR Technologies in Education Developments: Opportunities and challenges. In *2018 6th RSI international conference on robotics and mechatronics (IcRoM)* (pp. 482–487). IEEE. https://doi.org/10.1109/ICRoM.2018.8657615

Arnaldi, B., Guitton, P., & Moreau, G. (2018). *Virtual reality and augmented reality: Myths and realities* (1st ed.). Wiley-IEEE Press.

Bender, A. D., Filmer, H. L., Naughtin, C. K., & Dux, P. E. (2017). Dynamic, continuous multitasking training leads to task-specific improvements but does not transfer across action selection tasks. *Npj science of Learning*, *2*(1), 14. https://doi.org/10.1038/s41539-017-0015-4

Botch, T. L., Garcia, B. D., Choi, Y. B., Feffer, N., & Robertson, C. E. (2023). Active visual search in naturalistic environments reflects individual differences in classic visual search performance. *Scientific Reports*, *13*(1), 1–9. https://doi.org/10.1038/s41598-023-27896-7

Boucheix, J. M., Lowe, R. K., Putri, D. K., & Groff, J. (2013). Cueing animations: Dynamic signaling aids information extraction and comprehension. *Learning and Instruction*, *25*, 71–84. https://doi.org/10.1016/j.learninstruc.2012.11.005

Buttussi, F., & Chittaro, L. (2018). Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE Transactions on Visualization and Computer Graphics*, *24*(2), 1063–1076. https://doi.org/10.1109/TVCG.2017.2653117

Cheng, K.-H., & Tsai, C.-C. (2019). A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers & Education*, *140*, 103600. https://doi.org/10.1016/j.compedu.2019.103600

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, *19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Dahmani, L., & Bohbot, V. D. (2020). Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-62877-0

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, *41*(1), 10–32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Davis, R. O. (2018). The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis. *Educational Research Review*, *24*, 193–209. https://doi.org/10.1016/j.edurev.2018.05.002

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology*, *21*(6), 731–746. https://doi.org/10.1002/acp.1346

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, *20*(2), 111–122. https://doi.org/10.1016/j.learninstruc.2009.02.010

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2011). Attention cueing in an instructional animation: The role of presentation speed. *Computers in Human Behavior*, *27*(1), 41–45. https://doi.org/10.1016/j.chb.2010.05.010

Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., & Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron*, *63*(1), 127–138. https://doi.org/10.1016/j.neuron.2009.06.005

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Friedman, A., Kohler, B., Gunalp, P., Boone, A. P., & Hegarty, M. (2020). A computerized spatial orientation test. *Behavior Research Methods*, *52*(2), 799–812. https://doi.org/10.3758/s13428-019-01277-3

Gardony, A. L., Brunyé, T. T., & Taylor, H. A. (2015). Navigational aids and spatial memory impairment: The role of divided attention. *Spatial Cognition and Computation*, *15*(4), 246–284. https://doi.org/10.1080/13875868.2015.1059432

Grogorick, S., Albuquerque, G., & Magnor, M. (2018). Comparing unobtrusive gaze guiding stimuli in head-mounted displays. In *Proceedings – international conference on image processing* (pp. 2805–2809). ICIP. https://doi.org/10.1109/ICIP.2018.8451784

Hazeltine, E., Teague, D., & Ivry, R. B. (2002). Simultaneous dual-task performance reveals parallel response selection after practice. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 527–545. https://doi.org/10.1037/0096-1523.28.3.527

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175–191. https://doi.org/10.1016/j.intell.2003.12.001

Hejtmánek, L., Oravcová, I., Motýl, J., Horáček, J., & Fajnerová, I. (2018). Spatial knowledge impairment after GPS guided navigation: Eye-tracking study in a virtual town. *International Journal of Human Computer Studies*, *116*, 15–24. https://doi.org/10.1016/j.ijhcs.2018.04.006

Hornof, A. J., Zhang, Y., & Halverson, T. (2010). Knowing where and when to look in a time-critical multimodal dual task. *Conference on Human Factors in Computing Systems – Proceedings*, *3*, 2103–2112. https://doi.org/10.1145/1753326.1753647

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning*, *22*(6), 392–404. https://doi.org/10.1111/j.1365-2729.2006.00180.x

Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology*, *28*(1), 74–82. https://doi.org/10.1016/j.jenvp.2007.09.002

Jiang, L., Xu, M., Liu, T., Qiao, M., & Wang, Z. (2018). DeepVS: A deep learning based video saliency prediction approach. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer vision – ECCV 2018* (pp. 625–642). Springer International Publishing. https://doi.org/10.1007/978-3-030-01264-9_37

Johnson-Glenberg, M. C. (2019). *The necessary nine: Design principles for embodied VR and active stem education*. Springer. https://doi.org/10.1007/978-981-13-8265-9_5

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology*, *107*(3), 724–739. https://doi.org/10.1037/edu0000013

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*, 1–18. https://doi.org/10.3389/fpsyg.2017.01997

Lee, K. M. (2004). Presence, explicated. *Communication Theory*, *14*(1), 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lin, L., & Atkinson, R. K. (2011). Using animations and visual cueing to support learning of scientific concepts and processes. *Computers and Education*, *56*(3), 650–658. https://doi.org/10.1016/j.compedu.2010.10.007

Lin, Y.-C., Chang, Y.-J., Hu, H.-N., Cheng, H.-T., Huang, C.-W., & Sun, M. (2017). Tell me where to look: Investigating ways for assisting focus in 360° video. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 2535–2545). ACM. https://doi.org/10.1145/3025453.3025757

Liu, R., Xu, X., Yang, H., Li, Z., & Huang, G. (2022). Impacts of cues on learning and attention in immersive 360-degree video: An eye-tracking study. *Frontiers in Psychology*, *12*, 1–11. https://doi.org/10.3389/fpsyg.2021.792069

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, *113*(4), 719–735. https://doi.org/10.1037/edu0000473

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, *66*(5), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, *72*, 276–285. https://doi.org/10.1016/j.chb.2017.02.066

Makransky, G., Mayer, R., Nøremølle, A., Cordoba, A. L., Wandall, J., & Bonde, M. (2020). Investigating the feasibility of using assessment and explanatory feedback in desktop virtual reality simulations. *Educational Technology Research and Development*, *68*(1), 293–317. https://doi.org/10.1007/s11423-019-09690-3

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, *33*, 937–958. https://doi.org/10.1007/s10648-020-09586-2

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, *60*(1), 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Marsh, T., & Smith, S. P. (2001). Guiding user navigation in virtual environments using awareness of virtual off-screen space. In V. Paelke & S. Volbracht (Eds.), *Proceedings of the workshop on guiding users through interactive experiences—Usability Centred design and evaluation of virtual 3D environments* (pp. 149–154). Shaker Verlag. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.4719

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part I Basic Mechanisms. *Psychological Review*, *104*(1), 3–65. https://doi.org/10.1037/0033-295X.104.1.3

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, *140*(8), 103603. https://doi.org/10.1016/j.compedu.2019.103603

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999-2009). *Computers and Education*, *56*(3), 769–780. https://doi.org/10.1016/j.compedu.2010.10.020

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition*, *21*(1), 59–68. https://doi.org/10.1016/j.concog.2011.08.010

Moritz, J., Meyerhoff, H. S., & Schwan, S. (2020). Control over spatial representation format enhances information extraction but prevents long-term learning. *Journal of Educational Psychology*, *112*(1), 148–165. https://doi.org/10.1037/edu0000364

Olmos-Raya, E., Ferreira-Cavalcanti, J., Contero, M., Castellanos, M. C., Giglioli, I. A. C., & Alcañiz, M. (2018). Mobile virtual reality as an educational platform: A pilot study on the impact of immersion and positive emotion induction in the learning process. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(6), 2045–2057. https://doi.org/10.29333/ejmste/85874

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, *110*(6), 785–797. https://doi.org/10.1037/edu0000241

Parong, J., & Mayer, R. E. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning*, *37*(1), 226–241. https://doi.org/10.1111/jcal.12482

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220–244. https://doi.org/10.1037//0033-2909.116.2.220

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, *17*, 19–36. https://doi.org/10.1016/j.edurev.2015.12.003

Rothe, S., & Hußmann, H. (2018). Guiding the viewer in cinematic virtual reality by diegetic cues. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 10850 LNCS* (pp. 101–117). Springer. https://doi.org/10.1007/978-3-319-95270-3_7

Schmidt, S., Bruder, G., & Steinicke, F. (2019). Effects of virtual agent and object representation on experiencing exhibited artifacts. *Computers & Graphics*, *83*, 1–10. https://doi.org/10.1016/j.cag.2019.06.002

Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educational Research Review*, *23*, 1–24. https://doi.org/10.1016/j.edurev.2017.11.001

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The cognitive-affective-social theory of learning in digital environments (CASTLE). *Educational Psychology Review*, *34*(1), 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schubert, T. (2003). The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift Für Medienpsychologie*, *15*, 69–71. https://doi.org/10.1026//1617-6383.15.2.69

Seok, K.-H., Kim, Y., Son, W., & Kim, Y. S. (2021). Using visual guides to reduce virtual reality sickness in first-person shooter games: Correlation analysis. *JMIR Serious Games*, *9*(3), 1–13. https://doi.org/10.2196/18020

Shi, L., Huang, F. C., Lopes, W., Matusik, W., & Luebke, D. (2017). Near-eye light field holographic rendering with spherical waves for wide field of view interactive 3D computer graphics. *ACM Transactions on Graphics*, *36*(6), 1–17. https://doi.org/10.1145/3130800.3130832

Smith, S. P., & Marsh, T. (2004). Evaluating design guidelines for reducing user disorientation in a desktop virtual environment. *Virtual Reality*, *8*(1), 55–62.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Speicher, M., Rosenberg, C., Degraen, D., Daiber, F., & Krüger, A. (2019). Exploring Visual Guidance in 360-degree Videos. In *TVX 2019 – proceedings of the 2019 ACM international conference on interactive experiences for TV and online video* (pp. 1–12). ACM. https://doi.org/10.1145/3317697.3323350

Spitzer, M. (2014). *Digitale Demenz*. Droemer Taschenbuch.

Stefanucci, J. K., Brickler, D., Finney, H. C., Wilson, E., Drew, T., & Creem-Regehr, S. H. (2022). Effects of simulated augmented reality cueing in a virtual navigation task. *Frontiers in Virtual Reality*, *3*, 1–16. https://doi.org/10.3389/frvir.2022.971310

Strobach, T., Frensch, P. A., Soutschek, A., & Schubert, T. (2012). Investigation on the improvement and transfer of dual-task coordination skills. *Psychological Research*, *76*(6), 794–811. https://doi.org/10.1007/s00426-011-0381-0

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

van der Haak, B. (2014). *Digital amnesia at work*. Kaspersky Lab https://media.kasperskydaily.com/wp-content/uploads/sites/89/2016/03/05232337/Digital-Amnesia-at-work-the-risks-and-rewards-of-forgetting-in-business-report.pdf

van Gog, T. (2021). The signaling (or cueing) principle in multimedia learning. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 221–230). Cambridge University Press. https://doi.org/10.1017/9781108894333.022

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Volkmann, T., Wessel, D., Caliebe, T. O., & Jochems, N. (2020). What you see isn't necessarily what you get. *Proceedings of Mensch Und Computer*, *60*(9), 119–128. https://doi.org/10.1145/3404983.3405590

Volkmann, T., Wessel, D., Jochems, N., & Franke, T. (2018). *German translation of the multimodal presence scale*. Universität zu Lübeck. https://doi.org/10.18420/muc2018-mci-0428

Wang, F., Li, W., Mayer, R. E., & Liu, H. (2018). Animated pedagogical agents as aids in multimedia learning: Effects on eye-fixations during learning and learning outcomes. *Journal of Educational Psychology*, *110*(2), 250–268. https://doi.org/10.1037/edu0000221

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Zhu, Y., Zhai, G., Min, X., & Zhou, J. (2020). The prediction of saliency map for head and eye movements in 360 degree images. *IEEE Transactions on Multimedia*, *22*(9), 2331–2344. https://doi.org/10.1109/TMM.2019.2957986

How to cite this article: Decker, D., & Merkt, M. (2024). Shake it or light it! The effects of cueing in desktop-VR learning environments on search time and learning. *Journal of Computer Assisted Learning*, *40*(3), 1201–1217. https://doi.org/10.1111/jcal.12945

Appendix C

Manuscript 3

Decker, D., & Merkt, M. (2025). Look at me! Can a pedagogical agent facilitate orientation and support learning in VR? *Educational Technology Research and Development.* https://doi.org/10.1007/s11423-025-10531-9

Author	Author Position	Scientific ideas %	Data generation %	Analysis and interpretation %	Paper writing %
Daniela Decker	1	80	100	90	80
Martin Merkt	2	20	0	10	20

Education Tech Research Dev
https://doi.org/10.1007/s11423-025-10531-9

RESEARCH ARTICLE

Look at me! Can a pedagogical agent facilitate orientation and support learning in VR?

Daniela Decker[1] · **Martin Merkt**[1]

Accepted: 14 June 2025

Abstract

Virtual reality (VR) is very promising for educational purposes but also presents learners with difficulties regarding orientation. Accordingly, VR environments should be designed to facilitate orientation, for example, by cueing. In a pre-registered laboratory experiment (between-subject design, 91 participants), we investigated the effects of a pedagogical agent cue compared to a light cue and a control condition without cues on search time, learning, mental representation, and perceived presence in a VR learning environment. Participants were tasked with locating tools in a virtual workshop environment, accompanied by a narration providing information about each tool. In the condition with the pedagogical agent cue, the agent was positioned close to the search objects and performed occasional gaze shifts to the object, whereas objects in the light cue condition were illuminated by a slightly reddish light. Both cueing methods significantly decreased search time but did neither affect learning outcomes nor the acquisition of the mental spatial representation of the learning environment. Additionally, the pedagogical agent cue reduced physical presence compared to the control condition and self-presence compared to both other conditions. In summary, these results imply that even with successful attention guidance, both types of cueing did not facilitate learning outcomes. It is an open question whether these effects generalize to larger VR environments and if different design choices regarding the pedagogical agent might influence presence positively, which might, in turn, lead to better learning outcomes.

Keywords Virtual reality learning · Cueing · Signaling · Dual- task paradigm · Presence

Introduction

Authentic settings such as research facilities for science education or engineering training workshops are environments that have been shown to significantly enhance learners' interest and engagement (Betz, 2018; Schüttler et al., 2021). Since accessing such real-world environments for educational purposes can be costly, dangerous or logistically challenging,

✉ Daniela Decker
Decker@die-bonn.de

[1] German Institute for Adult Education - Leibniz Centre for Lifelong Learning, Heinemannstraße 12-14, 53175 Bonn, Germany

using Virtual Reality (VR), i.e., a computer-generated environment which can be displayed on a head-mounted display (HMD), with a screen positioned directly in front of the eyes (Ardiny & Khanmirza, 2018; Buttussi & Chittaro, 2018), might present a valuable alternative by simulating these authentic settings in a controlled, immersive environment (Arnaldi et al., 2018; Fombona-Pascual et al., 2022).

Educational scenarios implemented in VR offer plenty of opportunities for hands-on practice (Arnaldi et al., 2018), especially since these systems have become more affordable (Parong & Mayer, 2018; Radianti et al., 2020). For instance, VR may enable the exploration of molecular interactions in chemistry (Fombona-Pascual et al., 2022), machinery assembly in engineering (Dolsak & Reich, 2022), and surgical training for medical professionals (Gazit et al., 2024; Landau et al., 2024). Overall, VR appears to be effective in supporting the acquisition of practical skills, such as motor or procedural tasks, although its effectiveness seems to depend on key features, such as immersion and presence (Tusher et al., 2024), the latter referring to the feeling of "being there" (Shu et al., 2019; Slater & Sanchez-Vives, 2016), a crucial prerequisite for learning in VR (Makransky & Petersen, 2021).

Despite these benefits, highly immersive VR environments also present challenges, particularly with regard to orientation (Grant & Magee, 1998; Makransky & Petersen, 2021). When presenting content in a 360° scene, creators need to ensure that users choose a viewpoint that represents the most relevant material (Bender, 2019), hence guiding them to orient themselves within the VR environment to avoid missing key elements. In this manuscript, the term "orientation" refers to participants selecting the most relevant viewpoint in the 360° scene. Various methods to enhance orientation within 360° scenes have been tested, for instance in VR videos (Beege et al., 2023) and desktop-VR (Decker & Merkt, 2024) with learning aspects, as well as in HMD-based VR without a focus on learning (Rothe & Hußmann, 2018).

Therefore, it is imperative to develop methods that improve orientation in immersive virtual learning environments without compromising representational fidelity, realism, and a high sense of presence, ensuring that VR's full potential as a powerful educational tool can be realized (Makransky & Petersen, 2021). Thus, the current experiment deals with cueing as a means of improving orientation in interactive VR learning environments. The signaling principle (see Sect."Cueing in VR learning environments") suggests that learners benefit from cues highlighting relevant parts of the learning material (Schneider et al., 2018), which may be particularly important in VR, where the immersive 360° environment can hinder the intuitive identification of the most relevant viewpoint. Cues such as arrows have been shown to be effective in assisting learners in finding relevant viewpoints in VR videos (Beege et al., 2023). However, those cues also may accentuate artificiality and reduce representational fidelity, a key factor influencing presence (Makransky & Petersen, 2021). To address this, the current experiment utilizes cueing, which involves elements that are an inherent part of the virtual scene or are perceived as such (Lange et al., 2020; Rothe & Hußmann, 2018).

Moreover, in our experiment, learners had to deal with a dual task (i.e., performing two tasks simultaneously) (Pashler, 1994), i.e., finding objects and listening to narrated learning content. Since previous research on dual tasks suggests that facilitating performance in one task can influence overall dual-task performance (Bender et al., 2017; Dux et al., 2009; Grinschgl et al., 2023; Hazeltine et al., 2002; Strobach et al., 2012), we aim to examine whether cueing that supports orientation in the search task frees up enough cognitive resources in the working memory to enhance overall performance, specifically by improving both search efficiency (i.e., shorter search times) and simultaneous learning

(i.e., encoding declarative knowledge in early long-term memory), as implied by previous research in desktop-VR (Decker & Merkt, 2024) (see Sect."Cueing in VR learning environments").

In the following sections, we will first provide an overview of key theories related to learning in VR, followed by an examination of the current research on cueing and pedagogical agents, with a focus on their potential to guide attention. Finally, we will explore the possible negative effects that cueing may entail.

Learning in VR learning environments

Focusing on the design of immersive learning environments, theories such as the Cognitive Affective Model of Immersive Learning (CAMIL; Makransky & Petersen, 2021) have emerged, building in part on the Cognitive Load Theory (CLT, Sweller et al., 1998). CLT highlights the limited capacity of working memory and distinguishes between three types of cognitive load: intrinsic load, which depends on task complexity and prior knowledge; germane load, referring to the effort invested in building and refining mental schemas; and extraneous load, which arises when cognitive resources are diverted by poorly designed materials or irrelevant elements (Klepsch et al., 2017; Sweller et al., 1998). Sharing the assumption of limited cognitive capacity, CAMIL extends this perspective by identifying three key technical factors that shape learning in VR: immersion, control, and representational fidelity (Makransky & Petersen, 2021). Immersion can be defined as a measure of the vividness of a system and the extent to which the system can exclude the external environment (Cummings & Bailenson, 2016). Further, control factors consist of the degree, immediacy, and nature of control the system allows (Witmer & Singer, 1998). Finally, representational fidelity is defined as the extent to which the representation corresponds to the real world (Dalgarno & Lee, 2010; Makransky & Petersen, 2021). While only control factors determine agency, characterized as the sense of creating and controlling one's actions (Johnson-Glenberg, 2019; Moore & Fletcher, 2012), all the technical factors mentioned influence presence. Presence is understood as the psychological feeling of 'being there' in a VR environment (Lee, 2004; Witmer & Singer, 1998). Lee (2004) suggests that presence can be experienced across three dimensions: physical, social, and self-presence. Physical presence pertains to interactions with objects and surroundings, social presence involves engagement with other social actors, and self-presence reflects individuals'perceptions of themselves within an environment.

Presence and agency are key psychological prerequisites for learning in VR, as they positively influence affective and cognitive factors like interest and motivation (Makransky & Petersen, 2021). Presence also may amplify the effect of immersion on flow, which in turn enhances learning transfer (Kang et al., 2022). Regarding agency, Johnson-Glenberg et al. (2021) conducted a study comparing 2D (PC) and 3D (VR) platforms, where participants experienced either high or low agency. High-agency participants directly manipulated virtual elements, while low-agency participants only observed pre-recorded gameplay. Results showed that high agency significantly improved learning in VR. However, low-agency participants in VR performed worse than all other groups (i.e., the PC conditions), suggesting that lacking agency in immersive environments may be especially detrimental to learning (Johnson-Glenberg et al., 2021).

High levels of immersion in VR might, however, also increase cognitive load (Makransky et al., 2019, 2021; Meyer et al., 2019; Parong & Mayer, 2021), possibly in part because learners are fully surrounded by the virtual environment. Therefore, they have to decide in

which direction they look and consequently only perceive their limited field of view (Zhu et al., 2020), which is the maximum angle that is momentarily visible and varies depending on the learners' position in the environment (Shi et al., 2017). Key areas may fall outside the learners' field of view in VR, making attention guidance more demanding than in materials with a predefined field of view (Grogorick et al., 2018; Makransky et al., 2021). Thus, the lack of effective guidance can lead to disorientation and navigational difficulties (Marsh & Smith, 2001; Smith & Marsh, 2004). Additionally, CAMIL also emphasizes that the need to search for relevant content might also increase extraneous cognitive load (Makransky & Petersen, 2021).

Cueing in VR learning environments

Since overcoming user disorientation is important for successful learning with VR, we will discuss cueing as a method to draw the learners'attention to the currently most important information (Alpizar et al., 2020; Boucheix et al., 2013; van Gog, 2021). According to the cueing (or signaling) principle, cues in learning materials successfully draw the learners' attention to relevant information, and consequently improve learning (Alpizar et al., 2020; Richter et al., 2016; Schneider et al., 2018; van Gog, 2021). For learning materials with a predefined field of view, cueing includes various options such as camera movements and zoom to the most relevant part of a video or animation (for other examples, see Schneider et al., 2018), which have already been successfully used for static or animated material (Boucheix et al., 2013; de Koning et al., 2011; Lin & Atkinson, 2011).

Various cueing techniques from traditional learning materials with a predefined field of view have been adapted for VR, with differing levels of effectiveness and intrusiveness. For example, forced rotation techniques like autopilot (i.e., forcibly directing the viewer's gaze to the target area) effectively direct gaze but can be obtrusive and may cause motion sickness due to sensory mismatches (Chang et al., 2020). In contrast, using a moving object to guide attention is also effective but less intrusive (for more examples, see Speicher et al., 2019). Moreover, cueing methods can be divided into diegetic and non-diegetic methods, where diegetic methods use elements within the original scene for attention guidance (e.g., sunlight, traffic lights) (Rothe & Hußmann, 2018) and non-diegetic methods add additional elements to the scene. An example of non-diegetic methods is the inclusion of arrows pointing in the intended direction, which could ease orientation in VR videos (Lin et al., 2017).

Regarding non-diegetic cues in VR videos, arrow cues and text cues enhanced learning outcomes (Liu et al., 2022), especially when they assisted learners to locate the relevant field of view (Beege et al., 2023). For interactive VR environments, research using text annotations as cues showed mixed results regarding learning outcomes (Albus et al., 2021; Zhang et al., 2023). Further, Albus and Seufert (2022) showed that green circles around relevant parts of a virtual animal (e.g., ears, hoofs) improved learning and decreased cognitive load in a desktop-VR learning environment. However, the cues in these experiments mainly guided learners' attention within objects that were already within the learners' field of view (Albus & Seufert, 2022; Albus et al., 2021). Further, whereas on the one hand, adding non-diegetic cues to the learning materials may be effective for highlighting relevant learning material, on the other hand, it might also emphasize that the environment is artificial, thus decreasing representational fidelity. Decreasing representational fidelity might in turn decrease presence and might thus hamper learning according to CAMIL (Makransky & Petersen, 2021).

Therefore, it may be worth exploring diegetic cues that are part of the scene and may thus be better suited to maintain representational fidelity. In this regard, light, movement, and sound cues have been shown to efficiently guide attention to relevant objects (Rothe & Hußmann, 2018). Extending the findings of Rothe and Hußmann (2018), we tested the effects of light and movement cues on a search task with simultaneously narrated learning content in a desktop-VR learning environment (Decker & Merkt, 2024). In two experiments, participants were tasked with locating and selecting objects in a desktop-VR learning environment while simultaneously listening to a narration providing educational content about each object. Experiment 1 utilized a balanced within-subject design, where each participant experienced all cueing conditions across different trials (i.e., each search task featured a cue distinct from the previous one). The results indicated that only the movement cue significantly reduced search times and that neither the movement nor the light cue had an effect on learning outcomes. Several explanations may account for these findings. First, regarding search times, the light cue may have been at a disadvantage, as movement may attract attention more automatically (Jiang et al., 2018) compared to a static light cue. Given the unpredictability of cue types from one task to the next in the within-subjects design, only the more automatic movement cue may have proved effective. Second, concerning learning outcomes, the narration for each search task substantially exceeded the average search time, allowing participants ample opportunity to focus on the content once the object was located, thus diminishing the potential benefits of cueing in the dual-task paradigm including a concurrent search and learning task. Building on these possible explanations, Experiment 2 adopted a between-subject design to enhance predictability; participants were randomly assigned to one of three distinct groups, each experiencing only one cue type across all search tasks. Additionally, the narration was significantly shortened to test whether cueing would be effective when participants needed to engage with both tasks concurrently. In this experiment, results showed that the light cue facilitated overall performance (i.e., search times and learning outcomes) (Decker & Merkt, 2024). This result is in line with previous research suggesting that facilitating performance in one task might influence overall performance in a dual task (Bender et al., 2017; Dux et al., 2009; Grinschgl et al., 2023; Hazeltine et al., 2002; Strobach et al., 2012). The movement cue, though it consistently reduced search times across both experiments, may have lacked impact on learning due to a reduction in the sense of presence (Decker & Merkt, 2024). Thus, further research is needed to identify cueing methods that do not reduce presence and direct attention naturally. Since Speicher et al. (2019) suggested that one of the most efficient ways to guide attention was adding a person to follow in a VR video, in the current experiment, we compare a pedagogical agent cue with the light cue, which was promising in the desktop-VR learning environment (Decker & Merkt, 2024).

Pedagogical agents as cues

Human-like virtual characters, also known as pedagogical agents (Apoki et al., 2022; Craig et al., 2002), might serve several purposes in learning material. In the following section, we will explore their potential to direct attention and convey social cues.

First, several studies implied that humans (especially human faces) naturally attract attention in conventional videos (Kizilcec et al., 2014; Sondermann et al., 2024) and that following a person in a 360° degree video is an efficient way to guide attention (Speicher et al., 2019). Thus, it seems reasonable to assume that a pedagogical agent might also help learners to orient themselves by guiding them to the relevant viewpoint in the learning environment. Further,

focusing on gestures, several studies suggested that pedagogical agents can guide attention by using pointing gestures (Khokhar & Borst, 2022), which often influenced learning positively (Davis, 2018; Li et al., 2019). Moreover, gaze is a highly relevant source of information (McKay et al., 2021). People tend to align their attention to the same object as the person they are interacting with (Frischen et al., 2007). Thus, using gaze to direct attention to an object of mutual interest appears feasible. Moreover, a meta-analysis of McKay et al. (2021) implied that this so-called gaze-cueing effect appeared equally for real, schematic, and computer-generated faces (McKay et al., 2021). Therefore, we assume that the gaze of a pedagogical agent should provide enough information to direct attention. However, previous research on the effects of a pedagogical agent directing attention through gaze has shown heterogeneous results regarding its impact on learning (Li et al., 2023; Pejsa et al., 2015).

Second, using pedagogical agents to provide social cues in learning materials may be beneficial for learning (Bandura, 1977; Mayer, 2014; Schneider et al., 2022), as such cues can elicit a social response (Li et al., 2022), thereby enhancing (para-)social and metacognitive processes. These processes, in turn, can influence information selection and processing, as well as the construction of mental models in long-term memory.

In summary, the question remains whether a pedagogical agent could effectively direct attention in a 360° environment, and how such guidance impacts learning in VR. Additionally, we aim to investigate whether the agent's image can introduce social cues beyond the voice, which remained consistent across other conditions, into the learning environment.

Potential negative effects of cueing on the formation of a mental representation

Since a growing amount of research is concerned with adverse effects of digital support on human abilities (for an overview, see Clinch et al., 2021), investigating whether cueing negatively affects the formation of a mental spatial representation seems vital for the further development of new cueing methods in VR. For instance, navigation using GPS tools is an important area in which many people rely on digital support, even though current research implies detrimental effect on the construction of mental spatial representations (Dahmani & Bohbot, 2020; Gardony et al., 2015; Hejtmánek et al., 2018; Ishikawa et al., 2008).

Potential reasons for adverse effects of technology support on the construction of mental representations might be inappropriate encoding due to a lack of practice (van der Haak, 2014) or due to cognitive offloading, which is defined as the externalization of cognitive processes (Grinschgl et al., 2023). A recent review by Skulmowski (2023) emphasized that even though offloading behavior can be helpful in daily tasks, it might have negative effects on learning. Moreover, recent research has also shown that supporting learners in a practice phase by allowing offloading behavior (i.e., manipulating a digital representation according to the requirements of the task) enhanced immediate task performance; however, participants couldn't maintain their performance without this support (Moritz et al., 2020). Thus, we aim to investigate whether cueing support in a VR learning environment has detrimental effects on the construction of mental spatial representations.

The current experiment

In the current experiment, we compared the effects of cueing provided by a pedagogical agent to a light cue and a control condition without cueing. The dependent variables were search times, learning outcomes, mental spatial representation, and presence. Findings

from our previous research (Decker & Merkt, 2024) on cueing in a desktop-VR learning environment indicated that light cues reduced search times and improved learning outcomes in a between-subject design in that participants completed all search trials with the same cue. However, since according to CAMIL (Makransky & Petersen, 2021), an additional increase in presence could further improve learning, we decided to use a pedagogical agent cue in this experiment to increase social presence. In this context, cueing provided by a pedagogical agent might be effective, as learners are expected to direct their field of view toward the human-like agent, since humans naturally attract attention (Kizilcec et al., 2014).

Further, previous research suggested that performing two tasks simultaneously (i.e., performing search tasks and simultaneously listening to learning materials) can be particularly challenging for learners (Bender et al., 2017; Hornof et al., 2010; Meyer & Kieras, 1997; Pashler, 1994). Facilitating the successful completion of one of the two tasks that participants are expected to perform simultaneously may improve overall performance. Approaches tested include, for instance, training or cognitive offloading (e.g., providing opportunities to externalize mental processes, for instance by writing them down) (Bender et al., 2017; Dux et al., 2009; Grinschgl et al., 2023; Hazeltine et al., 2002; Strobach et al., 2012). The mechanism underlying this approach is based on the premise that when cognitive load is reduced in one task, additional resources may become available for the other one (Grinschgl et al., 2023). In this study, we aim to test whether cueing might also serve as an effective strategy to achieve this redistribution by guiding attention in a way that minimizes the demands of the search task and potentially enhances learning outcomes in the simultaneous learning task.

Finally, since several studies implied detrimental effects on memory when participants overly rely on technical support (for an overview, see Clinch et al., 2021), we investigate whether cueing may have detrimental effects on learners' mental spatial representations. Moreover, we investigate how these different cueing methods affect presence, which is considered to be a prerequisite for successful learning in VR (Makransky & Petersen, 2021). Building on these considerations, we preregistered the following hypotheses on OSF.[1]

Hypotheses

Hypothesis 1 *The search task is solved faster if the object is highlighted by either the pedagogical agent cue* (*Hypothesis* 1*a*) *or the light cue* (*Hypothesis* 1*b*).

Hypothesis 2 *Learning outcomes will be better when the search task in the learning environment is supported by either the pedagogical agent cue* (*Hypothesis* 2*a*) *or the light cue* (*Hypothesis* 2*b*).

Hypothesis 3 *The mental spatial representation of the learning environment will be less accurate when the search task was supported by either the pedagogical agent cue* (*Hypothesis* 3*a*) *or the light cue* (*Hypothesis* 3*b*) (*i.e., participants will remember fewer object positions accurately*).

[1] https://osf.io/7ufrh/?view_only=28ae330c9bb344b191b66428d1f50b83

Hypothesis 4 *Presence* (*Presence scale*, *see Makransky et al.*, 2017) *will be higher when the search task in the learning environment is supported by the pedagogical agent cue compared to the no cueing condition* (*Hypothesis* 4*a*) *and the light cueing condition* (*Hypothesis* 4*b*).

Method

Participants and design

The experiment was conducted with 91 participants, none of whom had to be excluded due to the pre-defined exclusion criteria (i.e., not completing the study, insufficient knowledge of the German language, missing values in the search times, severe motion sickness, incorrect answer to the attention check). This resulted in a sample of 91 German-speaking participants (46 male, 45 female). Their mean age was 37.44 years ($SD = 14.28$). In this experiment, the independent variable was cueing with the factor levels pedagogical agent cue, light cue, and no cue. The dependent variables were search times, learning, mental spatial representation, and presence. Participants were randomly assigned to one of the three between-subjects conditions and received 10 € for their participation.

Learning materials: VR learning environment

The VR learning environment represented a virtual wood workshop which was presented on the Pico Neo 2 Eye, a stand-alone VR headset with a 101° field of view, 6 DoF tracking and a refresh rate of 75 Hz. In this learning environment, participants used a virtual laser pointer to locate and select twelve typical workshop items, such as a level, various tools, and a fire extinguisher. While they were searching, they listened to narrated factual information[2] about each object, which was delivered either by the pedagogical agent or an off-screen voice in the light cue condition and the condition without cues. The objects were distributed evenly across the environment and the laser pointer was controlled with the right VR controller. At the beginning of each of the twelve trials (eleven plus one practice trial), a large user interface (UI) centered in the participants' field of view showed a picture of the object and a prompt such as "Please find the fire extinguisher". After sufficient reading time the large UI disappeared, and a smaller version appeared on a virtual pinboard hanging on a wall in the workshop. A human male voice narrated the spoken learning content while the participants were searching. The narration took between 8 and 16 s for each object. The selection of the right item was confirmed with a green check mark appearing in the center of the field of view. After participants selected the object, the search trial proceeded until the spoken learning content was finished. If the participants could not find the object within 300 s, the trial was registered as incomplete. After a trial was either completed or registered as incomplete, the learning environment reloaded the scene, which was indicated by a short black overlay, and the next search trial started. The learning environment (see Fig. 1) was created using Unity (Version 2019.4.13f1).

[2] The full content of the narrated factual information is provided in Appendix B.

In the light cue condition, the cued object was illuminated with a reddish light imitating light coming from an invisible projector placed above the object. In the pedagogical agent cue condition, the pedagogical agent was positioned in an unoccupied area in close proximity to the search object, which might guide learners to the relevant field of view. The full-body pedagogical agent displayed conversational gestures without deictic functions (Davis, 2018) and mimic as he presented the learning content. In the beginning of each trial, the agent looked in the direction of the learners who stood in a predefined area in the middle of the virtual room. After a short time, the agent then performed occasional gaze shifts to the search object, which might have given learners an additional hint about the exact position of the object. After a few sec, the agent turned its gaze back to the learners. Regarding the design, the pedagogical agent was portrayed to be an adult male and wore brown dungarees. The agent's clothing design was based on the recommendations for workwear of the Employer's Liability Insurance Association for Wood and Metal in Germany (BGHM, 2019) and was evaluated to be credible and fitting to the theme in a previous pilot study (Decker & Merkt, 2023). The pedagogical agent was designed with Character Creator 3.3 and animated with iClone 7.8 (see Fig. 2).

Measures & instruments

Prior knowledge & interest Before engaging with the virtual reality (VR) learning environment, participants' self-assessed levels of prior knowledge and interest in the contents were measured with a 7-point Likert scales ranging from 1 (indicating almost no prior

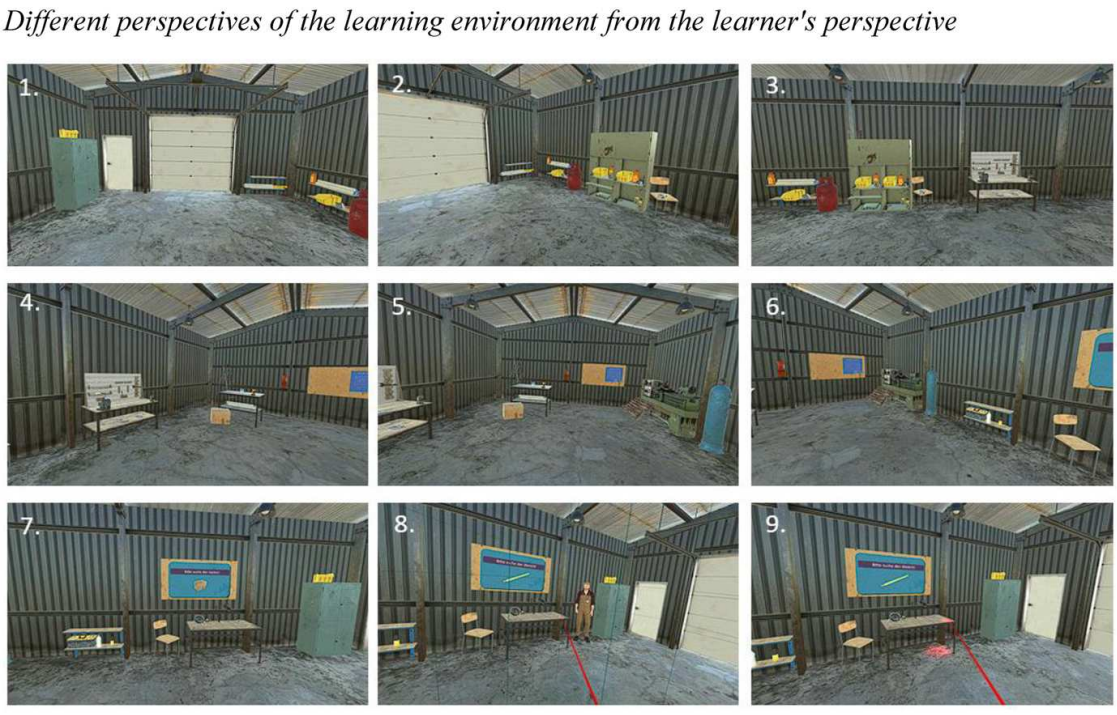


Fig. 1 Different perspectives of the learning environment from the learner's perspective. Note. 1–7. Step-by-step sections of the VR learning environment in a clockwise direction including the virtual pinboard; (8) Example for the placement of the pedagogical agent; (9) Example of an object with light cue

Pedagogical agent design

Note. The pedagogical agent was placed in close proximity to the search objects.

Fig. 2 Pedagogical agent design

knowledge/interest) to 7 (reflecting a lot of prior knowledge/very great interest). There were three items for each measure, covering topics like the usage of various tools (e.g. hammers, drills), weather protection of wood, and fire safety in workshops. Cronbach's alpha was 0.79 for the prior knowledge items and 0.82 for the interest items.

Additionally, participants' experiences with head-mounted display VR and desktop-VR as well as their experience with VR learning environments were assessed using three items that had to be rated on 7-point Likert scales ranging from 1 (indicating almost no experience) to 7 (indicating a lot of experience). Cronbach's alpha was 0.74 for the three VR experience items.

Search time During the search trials, the learning environment recorded the search times for each object, defined by the seconds elapsed from the start of the trial until the selection of the correct object with the virtual laser pointer.

Mental spatial representation To assess learners' mental spatial representation, a mental spatial representation test was created including the twelve search objects from the learning environment. This test environment resembled the original learning environment, except that red question marks were inserted as placeholders instead of the search objects and there was no narration. During this test, participants were prompted to select the placeholder that represented the position of each search object in the original learning environment. If participants could not find the correct object on the first attempt, they could try again until a maximum of 30 s had elapsed. The mental spatial representation test recorded the time it took participants to select the correct question mark and the number of times participants selected the wrong question mark, i.e. the number of mistakes. See Fig. 3 for an illustration.

Mental Spatial Representation Test

Note. Participants were asked which placeholder (red question mark) represented the search object in question in the learning environment.

Fig. 3 Mental Spatial Representation Test

Knowledge test Learning outcomes were measured with 22 open-ended questions (i.e., two questions per search object). No knowledge questions were posed about the object in the practice trial. The responses were scored independently by two raters, with all Cohen's Kappa coefficients ≥ 0.80, indicating strong inter-rater agreement. Each correct answer was awarded one point, resulting in a maximum score of 22 points. For correct but incomplete answers, partial points were assigned. The Cronbach's alpha coefficient for the knowledge items was 0.73.

Presence

We assessed physical presence, self-presence, and social presence utilizing a set of 5 items each (i.e., a total of 15 items for the complete presence assessments) developed by Makransky et al. (2017). For physical presence, participants responded to items such as "I was completely captivated by the virtual world" (Cronbach's alpha = 0.87). Self-presence was measured with items including "I felt like my real hand was inside of the virtual environment"(Cronbach's alpha = 0.87), while social presence was assessed with statements like "I felt like I was in the presence of another person in the virtual environment"(Cronbach's alpha = 0.88). These items were rated on a 5-point Likert scale ranging from 1 (does not apply at all) to 5 (fully applies). In this experiment we used the German version of the scale, translated by Volkmann et al. (2018).

Procedure of the experiment

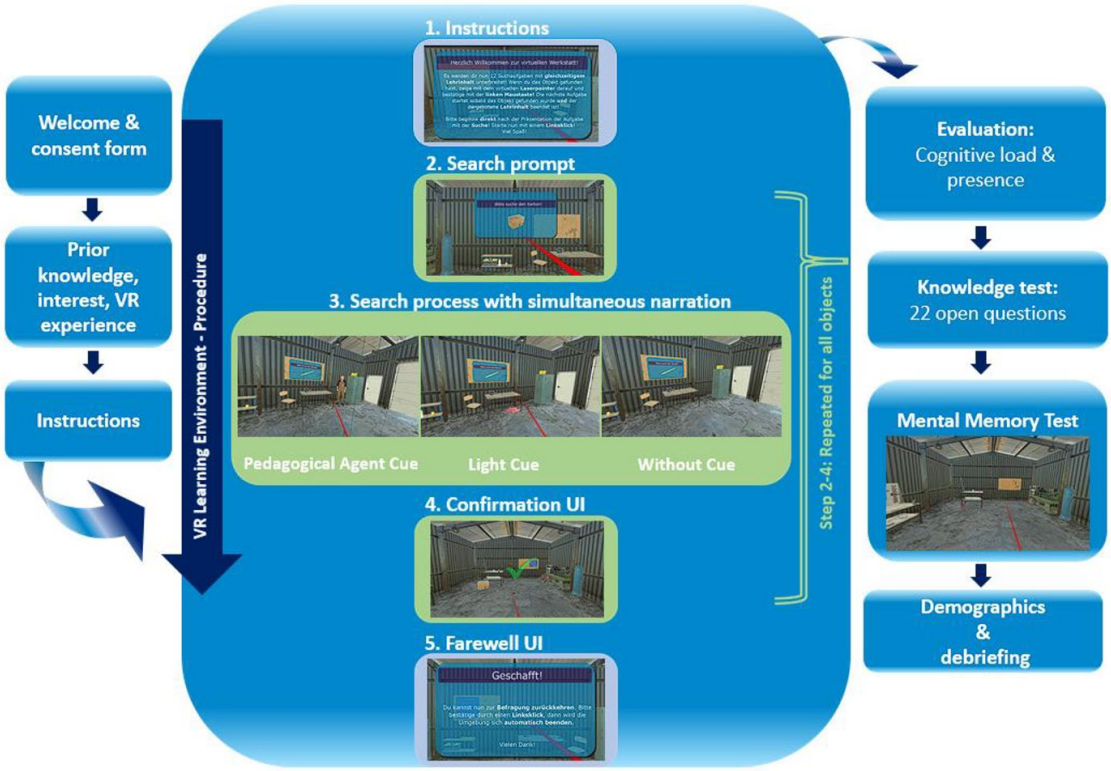


Fig. 4 Procedure of the experiment

Cognitive load

We employed the cognitive load scale developed by Andersen and Makransky (2021). This instrument was specifically adapted for virtual reality environments. The scale measures cognitive load using 11-point Likert scales ranging from 0 (does not apply at all) to 10 (fully applies). In addition to measuring germane cognitive load (4 items, such as "The virtual environment really enhanced my understanding of the procedures covered", Cronbach's alpha = .92) and intrinsic cognitive load (3 items, such as "The virtual environment covered procedures that I perceived as very complex", Cronbach's alpha = .94), this scale distinguishes between three types of extraneous cognitive load. These include environmental extraneous cognitive load (4 items, such as "The virtual environment was full of irrelevant content", Cronbach's alpha = .72), extraneous cognitive load for instruction (3 items, such as "The explanations used in the virtual environment were full of unclear content", Cronbach's alpha = .53), and extraneous cognitive load for interaction (4 items, such as "The interaction technique used in the virtual environment made it harder to learn", Cronbach's alpha = .57).

Procedure

A maximum of two participants per timeslot were seated in front of a laptop in the laboratory and asked to sign the consent form. Afterwards, they were randomly assigned to one of the three experimental conditions (i.e. pedagogical agent cue, light cue, or no cue). Next, prior knowledge, general interest in the learning content, and VR experience was recorded. The participants then completed the learning environment (see Sect."Learning materials: VR learning environment"for a detailed description) in the version that corresponded to their randomly assigned condition. Participants stood while engaging with the learning environment, which was delivered through the VR headset and lasted approximately five minutes, including initial instructions and a practice trial. After participants completed the learning environment, they were again seated in front of the laptop and evaluated their learning experience concerning cognitive load and presence. Next, the knowledge test was administered. Afterwards, the mental spatial representation was measured using the VR Headset, which took a maximum of six minutes (i.e., 30 s maximum for each object). Finally, participants were asked to provide their demographic data and the debriefing took place. See Fig. 4 for an illustration of the procedure.

Table 1 Overall means and standard deviations

Variable	Cueing method			Overall
	PA	Light	None	
Prior interest	3.09 (1.54)	3.29 (1.62)	3.02 (1.30)	3.14 (1.48)
Prior knowledge	2.74 (1.28)	3.30 (1.22)	3.09 (1.37)	3.04 (1.29)
VR experience	2.33 (1.37)	2.69 (1.54)	2.19 (1.34)	2.41 (1.42)
Search time	11.87 (5.05)	9.28 (5.62)	15.94 (6.78)	12.38 (6.44)
Knowledge test	5.91 (3.06)	7.06 (3.27)	7.02 (3.73)	6.68 (3.38)
Mental representation speed—T_{50}	4.06 (1.37)	4.04 (1.03)	4.21 (2.50)	4.11 (1.74)
Mental representation first hit success	6.54 (1.77)	6.87 (1.77)	7.32 (1.81)	6.91 (1.79)
Mental representation total mistakes	1.99 (1.64)	1.78 (1.52)	1.64 (2.12)	1.80 (1.77)
Social presence	1.77 (0.74)	2.14 (0.98)	2.10 (1.05)	2.01 (0.94)
Physical presence	3.27 (0.91)	3.78 (0.62)	3.88 (0.89)	3.65 (0.85)
Self presence	2.60 (0.77)	3.24 (0.98)	3.21 (1.00)	3.02 (0.96)
Intrinsic cognitive load	3.30 (2.77)	3.48 (2.05)	3.17 (2.96)	3.32 (2.59)
Environmental extraneous cognitive load	2.22 (1.87)	1.70 (1.71)	2.13 (1.76)	2.01 (1.77)
Extraneous cognitive load for instruction	2.06 (1.53)	1.82 (1.68)	2.12 (1.86)	2.00 (1.68)
Extraneous cognitive load for interaction	2.34 (1.81)	1.37 (1.22)	1.91 (1.73)	1.86 (1.63)
Germane cognitive load	2.90 (2.55)	3.94 (1.94)	3.47 (2.82)	3.45 (2.47)
N	29	31	31	91

Results

Data were analyzed using ANOVAs with cueing as a between-subject factor (pedagogical agent cue, light cue, no cue) for each of the dependent variables. The data and the analysis script are available on OSF.[3] All descriptive data can be seen in Table 1.[4]

Prior knowledge, interest, VR experience, and demographics

There were no differences between the three groups concerning learners' self-reported prior knowledge ($M = 3.05$, $SD = 1.30$), $F(2, 88) = 1.46$, $p = 0.237$, $\eta_p^2 = .03$, self-reported interest about the contents ($M = 3.14$, $SD = 1.48$), $F < 1$, or self-reported VR experience ($M = 2.41$, $SD = 1.42$), $F(2, 88) = 1.00$, $p = .372$, $\eta_p^2 = .02$.

Further, neither the proportion of gender X^2 $(2, N = 91) = 5.78$, $p = .055$, nor the proportion of highest educational qualifications, X^2 $(8, N = 91) = 5.97$, $p = 0.651$, differed between the groups. Regarding age, there was no effects of cueing, $F < 1$. Thus, we can assume that the groups were comparable.

Search time

Corroborating Hypotheses 1a and 1b, the ANOVA for search time revealed a significant effect of cueing, $F(2, 88) = 10.13$, $p < .001$, $\eta_p^2 = .19$. Post-hoc tests with Bonferroni adjustment showed faster search times for the pedagogical agent cue ($M = 11.87$, $SD = 5.05$) compared to the condition without cueing ($M = 15.94$, $SD = 6.78$), $p = .026$, and also faster search times for the light cue ($M = 9.28$, $SD = 5.62$) compared to the condition without cueing, $p < .001$. The pedagogical agent cue and the light cue did not differ, $p = 0.272$.

Learning outcomes

In contrast to Hypotheses 2a and 2b, the ANOVA revealed no significant effect of cueing for the learning outcomes ($M = 6.68$, $SD = 3.38$), $F(2, 88) = 1.10$, $p = 0.336$, $\eta_p^2 = .02$.

Mental spatial representation

Concerning mental spatial representation, three different analyses were conducted: one examining the times participants correctly picked the correct placeholder without committing mistakes, another analyzing the total number of mistakes made across trials, and a third focusing on the top 50% of search times until the correct placeholder was selected.

The first analysis regarding the number of times participants successfully picked the correct placeholder without mistakes did not reveal any effect of cueing, $F(2, 88) = 1.54$, $p = 0.220$, $\eta_p^2 = 0.03$.

[3] https://osf.io/u6z8k/?view_only=dcdd029041624a8e954b1303909b38ff

[4] The full list of items in the instruments used in this study is provided in Appendix B. Additionally, a MANOVA encompassing the dependent variables has been included in Appendix A.

The second analysis regarding the total number of mistakes participants made in the selection of the correct object placeholder also did not reveal an effect of cueing, $F < 1$.

For the third analysis, the search times until the correct placeholder was selected were analyzed. Since several participants did not identify the correct placeholder for some objects within 30 s, there were missing values, which could distort the calculation of the means. Thus, we applied the T_{50}—Method (Hooge & Camps, 2013) and included only the top 50% of each participants search times for the correct placeholder. This analysis also did not reveal an effect of cueing, $F < 1$. Thus, none of the analyses supported Hypotheses 3a and 3b.

Presence

Concerning Hypothesis 4, the ANOVAs revealed no significant effect of cueing on social presence, $F(2, 88) = 1.42$, $p = 0.246$, $\eta_p^2 = .03$, but significant effects on physical presence, $F(2,88) = 4.74$, $p = .011$, $\eta_p^2 = 0.10$, and self-presence, $F(2, 88) = 4.47$, $p = .014$, $\eta_p^2 = .09$. Regarding physical presence, Bonferroni-adjusted post-hoc tests showed lower physical presence for the pedagogical agent cueing condition ($M = 3.27$, $SD = 0.91$) compared to the condition without cueing ($M = 3.88$, $SD = 0.89$), $p = .015$. The light cue condition did not significantly differ from the pedagogical agent cue condition ($M = 3.78$, $SD = 0.62$), $p = 0.055$, and the condition without cueing, $p > 0.999$. Concerning self-presence, Bonferroni-adjusted post-hoc tests also showed lower self-presence for the pedagogical agent cueing condition ($M = 2.60$, $SD = 0.77$) compared to the condition without cueing ($M = 3.21$, $SD = 1.00$), $p = .039$, and the light cueing condition ($M = 3.24$, $SD = 0.98$), $p = .027$. The light cueing condition and the condition without cueing did not differ, $p > .999$.

Cognitive load

Concerning the explorative evaluation of cognitive load, the ANOVAs did not reveal any significant effects. Specifically, there were no observed effects of cueing on intrinsic cognitive load ($M = 3.32$, $SD = 2.59$), $F < 1$, germane cognitive load ($M = 3.45$, $SD = 2.47$), $F(2, 88) = 1.34$, $p = 0.268$, $\eta_p^2 = .03$, extraneous cognitive load for the environmental scale ($M = 2.01$, $SD = 1.77$), $F < 1$, extraneous cognitive load for instruction ($M = 2.00$, $SD = 1.68$), $F < 1$, or extraneous cognitive load for interaction ($M = 1.86$, $SD = 1.63$), $F(2, 88) = 2.74$, $p = .070$, $\eta_p^2 = .06$.

Discussion

Interpretation of the results

In this experiment, we investigated the effects of a pedagogical agent cue—compared to a light cue and to a control group without cueing—on search times, learning outcomes, mental spatial representation, and presence. Regarding Hypotheses 1a and 1b, light cues and the cueing provided by the pedagogical agent both reduced search times compared to the control group. Thus, both cueing methods guided attention equally successful with large

effect sizes. However, neither the light nor the pedagogical agent cue improved learning outcomes compared to the control group (Hypotheses 2a & 2b).

The observation that the light cue did not support learning was unexpected, as it successfully supported learning in a similar task in desktop-VR (Decker & Merkt, 2024). This observation might indicate that immersive VR is more challenging, possibly because physically moving around with the whole body to perceive the environment might consume more resources compared to steering a desktop-VR using mouse and keyboard. Moreover, using a VR headset might still be a new and overwhelming experience for many participants compared to using a conventional PC set up, and thus leads to inferior learning. Therefore, it is noteworthy that the overall learning results across the two comparable conditions (light cueing and no cueing) in this experiment were lower ($M = 7.04$, $SD = 3.48$) compared to our previous study ($M = 9.07$, $SD = 3.67$),[5] even though physical and self-presence in this experiment far surpassed that of our previous research.[6] However, these comparisons should be interpreted with caution, because beyond the difference in media (immersive VR vs. desktop-VR), there were other differences between these experiments. In the desktop-VR, the mini user interface (which appeared after the large one disappeared) was placed in the upper right corner of the screen, whereas in the HMD VR, it was on a virtual pinboard. Additionally, desktop-VR participants started each task from the same position, while participants in this experiment began each new task from where the last one ended to reduce motion sickness from forced repositioning. Moreover, while the previous experiment was conducted online, the current study took place in a laboratory setting, introducing additional factors beyond the difference in medium. One such factor could be differences in participant populations: the online sample was recruited via the panel provider Prolific, where registered participants are likely more experienced with study platforms and familiar with the technical systems used in research. Moreover, a comparison of self-assessed prior knowledge and VR experience between the experiments revealed that participants in this study reported significantly greater prior knowledge[7] of the learning content, whereas those in our previous research indicated significantly more experience with VR.[8] However, these comparisons, like those of learning outcomes, have interpretational limitations as they stem from different experiments. Nevertheless, systematically investigating how these variables influence learning outcomes—and whether, for example, pre-training (Delgado & Mayer, 2025) could help balance differences in prior VR experience—remains an interesting avenue for future research.

Aside from these differences, the tasks and learning content remained identical, and the gender distribution was comparable between the studies.[9] Thus, more systematic research might be necessary to test why cueing could not relieve enough cognitive resources to support learning in immersive VR, even though supporting the orientation seemed successful. Nevertheless, this result is in line with previous research by Makransky et al. (2019) stating that immersive VR decreased learning compared to desktop-VR. Another possible

[5] For comparisons between this experiment and our previous research, we analyzed aggregated means of comparable conditions (light cues and no cues) using one-sample t-tests and a chi-square test for gender: Learning outcomes: $t(61) = -4.59$, $p < .001$.

[6] Physical presence: $t(61) = 12.11$, $p < .001$/Self-Presence: $t(61) = 11.34$, $p < .001$.

[7] Prior knowledge: $t(61) = 4.24$, $p < .001$.

[8] VR experience: $t(61) = -3.69$, $p < .001$.

[9] Gender: $\chi 2(2, n = 168) = 1.16$, $p = .561$. The sample comprises the combined light cue and no-cue conditions from both experiments, including 106 participants from the previous study.

explanation relates to the types of memory engaged in the experiment. First, we hypothesized that cueing might reduce cognitive load in working memory, thereby facilitating the construction of schemas in early long-term memory. Engagement of the early long-term memory for the tests in this experiment can reasonably be assumed for two reasons: First, after the learning phase, participants evaluated their learning experience (e.g., presence and cognitive load), meaning that a significant amount of time passed between exposure to the VR environment and the knowledge test—far exceeding the few seconds typically associated with working memory (Baddeley, 1992). Second, the learning content, including the narration and spatial object positions, exceeded the assumed capacity of working memory, with the narration, for example, being too long to be rehearsed within the phonological loop (Baddeley, 1992). However, since CAMIL (Makransky & Petersen, 2021) suggests that highly immersive VR environments may be particularly beneficial for procedural knowledge, the fact that our study tested declarative memory, specifically the retention of semantic factual knowledge and spatial information, may have led to a misalignment between the strengths of the VR medium and the type of knowledge being tested, potentially influencing the outcomes negatively.

An additional factor to consider is cognitive load. Even though both cueing methods successfully reduced search times, participants did not rate their extraneous cognitive load to be lower in the cueing conditions. Thus, cueing did not free up cognitive resources towards a deeper elaboration of the learning materials, which might have enhanced overall performance in a dual task (Grinschgl et al., 2023). Managing two tasks simultaneously presents a significant challenge, often leading individuals to instinctively prioritize one of the tasks as the "primary task". In dual-task research, this tendency is a well-documented issue (Hazeltine et al., 2002). In our experiment, learners were informed that they would need to simultaneously search and listen, as the narrated content would be relevant for the knowledge test. However, due to the demands of the search task, which required the learners to actively locate objects in the virtual environment, it likely commanded more attention than the more passive task of listening to the learning materials. This may have decreased attention to the learning materials and dampened the effects of cueing on the learning outcomes. This explanation aligns with Jackson et al., (2023a, 2023b), who found that in a dual-task setup, where participants searched for hidden objects in a simulated environment while memorizing words, performance in word memorization decreased in comparison to a single task, as participants prioritized the search task over the memory task.

Potential explanations why cueing provided by the pedagogical agent did not support learning might also be related to the observations regarding presence (Hypothesis 4). Not only did the pedagogical agent fail to enhance social presence, but it reduced physical and self-presence instead. Regarding social presence, it could be argued that the agent's voice, which came from off-screen in the light cue and the control condition, provided the same level of social presence as the image of the agent (also see Mayer et al., 2003). Moreover, because the learning environment was rather short in duration (approximately 5 min) and the engagement with the agent was limited to passive listening, establishing a social connection to the agent could have been challenging. Further, since the study was conducted in the laboratory, where one person from the research team was present (and in some cases a second participant), the likelihood to seek a social connection with a pedagogical agent was probably limited.

Moreover, according to CAMIL (Makransky & Petersen, 2021), the reduced physical and self-presence might be due to the representational fidelity of the pedagogical agent. First, at the beginning of each task, the pedagogical agent was repositioned. Participants did not see this repositioning, but the pedagogical agent changed the position during the

black overlay when the environment was reloaded, which might have been disruptive. Second, in line with recommendations by Dai et al. (2022), we aimed to create a highly embodied pedagogical agent featuring conversational gesturing, idle positions, breathing movements, and facial expressions while narrating. In this process, we may have created a pedagogical agent that was close to a real person, but did not resemble real persons sufficiently, so that an uncanny valley effect may have occurred (Mori et al., 2012). The uncanny valley effect states that the likability of artificial characters increases as a function of their human likeness, but that this likability drops shortly before reaching the level of real people. At this point, the digital characters reach a very high level of human likeness, but still show signs of artificiality, which makes many people feel unsettled. In addition, the results of Li et al. (2024) imply that on-screen instructors with higher realism do not improve learning, possibly because these instructors attract a lot of attention. These explanations should be addressed through further variations in agent design in future research.

Regarding Hypothesis 3, there were no detrimental effects of cueing on the participants' mental spatial representation, which is in contrast to our initial assumption, but corroborates results from Stefanucci et al. (2022) stating that cues provided inside the environment may not have the same detrimental effects as navigational aids provided on a separate device such as a GPS tool. In particular, the efficient use of external navigational aids requires users to look at their device, which distracts their attention away from their physical surroundings. Moreover, Gardony et al. (2013) point out that the impairment found in several studies on the use of navigation aids might be caused by a switch between first-person perspective navigation in the learning phase to a survey-like assessment of the spatial representation (e.g., looking from above) in the test phase. Thus, using the same perspective in the learning environment and in the mental spatial representation test in our experiment may have avoided negative effects of cueing on spatial representation.

Our findings offer several implications for the design of VR-based learning environments. While both cueing methods effectively guided attention, this did not lead to improved learning outcomes, suggesting that orientation support alone may not be sufficient in dual-task learning contexts. Designers should therefore be cautious in assuming that reducing orientation effort automatically frees up resources for deeper learning, if orientation and learning need to happen at the same time. Consequently, future research should explore how additional instructional strategies, for instance a more extended familiarization with the VR setup and environment, may help learners make more effective use of their available cognitive resources in immersive environments. Additionally, the pedagogical agent reduced physical and self-presence, underscoring the importance of carefully calibrating its level of realism. Pre-testing both the design elements and their animations is therefore advisable.

Limitations and directions for future research

This research is subject to some limitations. First, the virtual workshop consisted of a single medium-sized room in which most of the search tasks could be mastered by walking a few steps and turning around. The extent to which these results can be generalized to more complex and larger VR environments should therefore be the subject of further research. For the learning content, we selected factual information within a woodworking context, as pre-testing in our previous research (Decker & Merkt, 2023) showed this knowledge was

not widely known. Crucially, this context served as a proxy to examine a broader principle: whether cueing in a dual-task scenario—combining search and learning—can free up cognitive resources to support deeper learning. This principle is relevant across various contexts in which learners must orient themselves while engaging with instructional content. VR was particularly well-suited for studying this dynamic, as it provides a highly controlled environment for systematically presenting and monitoring complex tasks (Obourdin et al., 2024). Thus, while cueing was tested within a specific scenario, our goal was to evaluate it as a generalizable instructional principle for immersive learning environments.

However, the open-answer format and the substantial amount of information covered in a short time may have contributed to relatively high item difficulties.[10] Therefore, future research could explore this principle using content with greater prior familiarity or a more accessible format, such as multiple-choice questions. Future research will be essential to further explore its applicability in different VR learning contexts. Furthermore, several comparisons in this discussion stem from two distinct experiments—one conducted online and the other in a laboratory setting. While our findings suggest that VR displayed in an HMD may pose greater challenges than desktop-VR, which would align with prior findings by Makransky et al. (2019), future research should further test this conclusion to gain a deeper understanding of how different VR setups impact learning and cognitive load. Other influencing variables such as prior knowledge and VR experience, which differed significantly between the experiments, could also be controlled in future research.

As previously discussed, perceiving the search task as the primary task would also be a possible explanation for the lack of cueing effects on the learning outcomes because cueing cannot have a supporting effect on the learning task if the narration is deliberately not given sufficient attention. In future studies, this problem could be addressed, for example, by redesigning the learning tasks to include more active participation. Moreover, participants were aware they needed to search for objects, naturally shifting their field of view. It would be valuable to test these cueing methods in scenarios where no specific focus areas are indicated, requiring only general exploration. This could reveal whether cueing enhances comprehensive exploration.

The design and implementation of the pedagogical agent is also subject to limitations. Even though the design was assessed as suitable in a pilot study (Decker & Merkt, 2023), the agent had negative effects on physical and self- presence. As discussed above, the repositioning of the agent between tasks could have disrupted presence. Thus, further approaches could include a more natural transition between search tasks. For example, a more fluid repositioning in which the agent walks to the new position could reduce the negative effects on presence. Since Speicher et al. (2019) also suggested that following a person is a potential way to guide attention in a VR video, it seems reasonable to assume that following an agent might also successfully guide attention.

Further, the primary aim of this experiment was to test attention guidance using the position of a pedagogical agent, with gaze intended only as an additional hint. Although the overall gaze pattern (i.e., from the participant to the object and back to the participant) remained consistent, the gaze itself varied according to the object's position, sometimes requiring more or less head and upper body movement. Systematically exploring these variations in future experiments would be a valuable direction for further research. Finally,

[10] The exact item difficulties can be found in the Appendix B Table 3, listed alongside the respective items.

future research could also explore how different levels of interaction between learners and agent, such as establishing a dialogue, might affect presence and learning.

Conclusion

Since difficulties like user disorientation are a major challenge in VR learning environments, methods to assist orientation need to be developed and tested. In this experiment, both a pedagogical agent cue and light cues guided learners' attention and had no negative effects on the construction of a mental spatial representation of the environment. However, the cueing provided by the pedagogical agent negatively impacted physical and self-presence, and neither of these cueing methods increased the learning outcomes in a simultaneous learning task. Future research should address whether these results generalize to more complex virtual environments with more complex learning tasks.

Appendix A

Supplementary analyses and findings

In addition to the analyses described in the manuscript, we conducted a MANOVA to assess potential interactions across the dependent variables (Learning, Search Time, Physical Presence, and Self-Presence). The MANOVA results indicated a statistically significant effect of cueing on the combined dependent variables, *Wilks'* $\Lambda = 0.674$, $F(8, 170) = 4.63$, $p < 0.001$. However, post-hoc ANOVAs on each individual dependent variable revealed the same effects as discussed in the paper, indicating no additional interactions among the dependent variables.

Appendix B

Measurement instruments

See Tables 2, 3, 4, 5

Table 2 Prior Knowledge and Experience Items

Dimension	Question	Item
Prior knowledge	How would you rate your prior knowledge on the following topics?	• Tools (saws, etc.) and their use • Weather protection of wood (varnish, etc.) • Fire protection (fire classes and firefighting)
Prior interest	How interested are you in the following topics?	• Tools (saws, etc.) and their use • Weather protection of wood (varnish, etc.) • Fire protection (fire classes and firefighting)
VR experience	How much experience do you have with…	• Virtual Reality (with a VR headset)? • Desktop Virtual Reality (general PC games from a first-person perspective)? • Virtual Reality learning environments?

[a]We used the function colMeans() in R to compute item difficulty, which represents the proportion of correct responses for each item. Thus, we obtained the percentage of participants who answered each item correctly

Note. These items were answered on 7-point Likert scales from 1 to 7

Table 3 Spoken Narration and Knowledge Items

Search Object	Narration	Open knowledge items	Item difficulty[a]
Cardboard	The first shipping boxes were patented in 1903. The grammage of cardboard varies between 150 and 600 g/m^2. Each year, the paper industry in the EU produces approximately 40 million tons of paper and cardboard, of which around 22.7 million tons are produced by the German paper industry, with a recycled paper rate of about 76%	Trial object	
Vise	The vise is attached to the workbench and consists of a main body that holds the slide and spindle. The slide is the movable part and performs the clamping motion. Clamping force is generated by turning the spindle	What parts make up a vise?	.22
		How is the clamping force of a vise generated?	.34
Drill	Impact drills enable drilling into harder materials through mechanical impacts. Softer materials like wood should be drilled without impact to prevent tearing of the hole	When should drilling without impact be used?	.63
		Why should drilling without impact be preferred for metal and wood?	.63
Pencil	Contrary to early assumptions, a pencil's core is made of graphite, not lead, and can write up to 56 km. Its writing capability relies on the layered crystal structure of graphite, where the van der Waals bond between atomic layers is so weak that they can be easily worn off	Complete the sentence: A pencil lead is writable because the ___ bonds between the atomic layers in graphite are so weak that they can be worn away	.14
		How long (in km) can an average pencil draw a continuous line?	.11
Level	A level checks the vertical or horizontal alignment of an object. Its essential component is the bubble vial—a plastic or glass tube filled with liquid and a bubble of air or gas	What is the most important component of a level?	.44
		What is a bubble vial in a level, and what is it made of?	.55
Euro pallet	Euro pallets measure 120 × 80 cm and are made up of 9 wooden blocks and 11 boards, held together with 78 special nails. They weigh approximately 25 kg and can support a dynamic load of 1500 kg and a static load of 4000 kg	How many wooden blocks and boards does an Euro pallet consist of?	.22
		How much does an Euro pallet weigh?	.26
	Narration	Open knowledge items	Item difficulty
Fire Extinguisher	Water-based extinguishers are suitable for Class A fires, such as wood and coal. Foam extinguishers can also tackle Class B fires (wax and gasoline), and ABC powder extinguishers are effective against Class C fires, specifically gas fires	What extinguishing agents can a fire extinguisher contain?	.37
		What materials burn in a Class C fire?	.23

Table 3 (continued)

Search Object	Narration	Open knowledge items	Item difficulty[a]
Pliers	The earliest pliers were used to hold workpieces in fire. Pliers consist of two handles, a joint, and a head, and operate on the lever principle. In technical language, the joint is referred to as the"trade."	What were the first documented pliers used for?	.42
		What is the technical term for the joint of a pair of pliers?	.07
Hammer	A hammer consists of a head and a handle. The hammerhead has a face and a peen. The peen is the wedge-shaped part of the hammerhead, while the flat, convex, or round striking surface is called the face. The cross-section of a face can be circular, oval, or rectangular	What is the name of the flat, convex, or rounded striking surface of a hammer?	.01
		Which part of the hammerhead is referred to as the peen?	.26
Paint Can	Paint is a liquid or powder coating substance, with the binder as its essential component. If the binder is not in liquid form, a solvent is required, which must be colorless, must not negatively affect the binder, and should evaporate without residue	In what forms can paint be found?	.37
		What is the essential component of paint?	.05
Brush	Brushes use three types of bristles: natural, synthetic, or mixed. Water-based paints should be applied with synthetic bristles, as the water content in the paint would cause natural bristles to swell, disrupting the flow of paint	What are the three types of bristles available for brushes?	.14
		Which type of bristle is preferable for solvent-based paints?	.51
Blueprint	A blueprint is a set of construction drawings that represent all rooms of a building with complete structural and geometric information. Powerful CAD programs are used to create blueprints	Complete the sentence: A blueprint is understood as a set of ___ that represents all ___ of a structure	.26
		Which programs are used to create blueprints?	.44

Note. The narration was presented during the search tasks, while the Knowledge Test followed afterwards

Table 4 Multimodal Presence Scale from Makransky et al. (2017) used in this Experiment

Dimension	Item
Physical presence	• The virtual environment seemed real to me
	• I had a sense of acting in the virtual environment, rather than operating something from outside
	• My experience in the virtual environment seemed consistent with my experiences in the real world
	• While I was in the virtual environment, I had a sense of "being there."
	• I was completely captivated by the virtual world
Social presence	• I felt like I was in the presence of another person in the virtual environment
	• I felt that the people in the virtual environment were aware of my presence
	• The people in the virtual environment appeared to be sentient (conscious and alive) to me
	• During the simulation, there were times where the computer interface seemed to disappear, and I felt like I was working directly with another person
	• I had a sense that I was interacting with other people in the virtual environment, rather than a computer simulation
Self-presence	• I felt like my virtual embodiment was an extension of my real body within the virtual environment
	• When something happened to my virtual embodiment, it felt like it was happening to my real body
	• I felt like my real self was projected into the virtual environment through my virtual embodiment
	• I felt like my real hand was inside of the virtual environment
	• During the simulation, I felt like my virtual embodiment and my real body became one and the same

Note. The Items were answered on a five-point Likert-Scale from 1 to 5

Table 5 The Multidimensional Cognitive Load Scale for Virtual Environments from Andersen and Makransky (2021) used in this Experiment

Dimension	Item
Intrinsic CL	• The topic covered in the simulation was very complex
	• The simulation covered procedures that I perceived as very complex
	• The simulation covered concepts and definitions that I perceived as very complex
Extraneous CL instruction	• The instructions and/or explanations used in the simulation were very unclear
	• The instructions and/or explanations used in the simulation were, in terms of learning, very ineffective
	• The instructions and/or explanations used in the simulation were full of unclear content
Extraneous CL interaction	• The interaction technique used in the simulation was very unclear
	• The interaction technique used in the simulation was, in terms of learning, very ineffective
	• The interaction technique used in the simulation made it harder to learn
	• The interaction technique used in the simulation was difficult to master
Extraneous CL environment	• The elements in the virtual environment made the learning very unclear
	• The virtual environment was, in terms of learning, very ineffective
	• The virtual environment was full of irrelevant content
	• It was difficult to find the relevant learning information in the virtual environment
Germane CL	• The simulation really enhanced my understanding of the topics covered
	• The simulation really enhanced my knowledge and understanding of lab safety
	• The simulation really enhanced my understanding of the procedures covered
	• The simulation really enhanced my understanding of concepts and definitions

Note. The Items were answered on an 11-Point Likert-Scale from 0 to 10. The Term "lab safety" was customized to "workshop safety"

Funding Open Access funding enabled and organized by Projekt DEAL. There was no external funding for the experiment.

Data availability The data that support the findings of this study are openly available on OSF at: https://osf.io/u6z8k/?view_only=dcdd029041624a8e954b1303909b38ff

Declarations

Conflict of interest The authors have no conflict of interest.

Ethical approval The experiment was approved by the local ethics committee of the German Institute for Adult Education in Bonn (DIE-LEK 2022–002). All participants gave informed consent prior to the experiment.

References

Albus, P., & Seufert, T. (2022). Signaling in 360° desktop virtual reality influences learning outcome and cognitive load. *Frontiers in Education, 7*, 1–10. https://doi.org/10.3389/feduc.2022.916105

Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education, 166*, 104154. https://doi.org/10.1016/j.compedu.2021.104154

Alpizar, D., Adesope, O. O., & Wong, R. M. (2020). A meta-analysis of signaling principle in multimedia learning environments. *Educational Technology Research and Development, 68*(5), 2095–2119. https://doi.org/10.1007/s11423-020-09748-7

Andersen, M. S., & Makransky, G. (2021). The validation and further development of a multidimensional cognitive load scale for virtual environments. *Journal of Computer Assisted Learning, 37*(1), 183–196. https://doi.org/10.1111/jcal.12478

Apoki, U. C., Ali Hussein, A. M., Al-Chalabi, H. K. M., Badica, C., & Mocanu, M. L. (2022). The role of pedagogical agents in personalised adaptive learning: A review. *Sustainability (Switzerland), 14*(11), 1–17. https://doi.org/10.3390/su14116442

Ardiny, H., & Khanmirza, E. (2018). The Role of AR and VR technologies in education developments opportunities and challenges. *RSI International Conference on Robotics and Mechatronics (IcRoM)*. https://doi.org/10.1109/ICRoM.2018.8657615

Arnaldi, B., Guitton, P., & Moreau, G. (2018). *Virtual reality and augmented reality: Myths and realities* (1st ed.). Wiley-IEEE Press.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Bandura, A. (1977). *Social learning theory*. Prentice-Hall.

Beege, M., Nebel, S., Schneider, S., & Rey, G. D. (2023). The effect of microlevel and macrolevel signaling on learning with 360° videos. *Applied Cognitive Psychology, 37*(1), 232–246. https://doi.org/10.1002/acp.4023

Bender, A. D., Filmer, H. L., Naughtin, C. K., & Dux, P. E. (2017). Dynamic, continuous multitasking training leads to task-specific improvements but does not transfer across action selection tasks. *Npj Science of Learning*. https://doi.org/10.1038/s41539-017-0015-4

Bender, S. (2019). Headset attentional synchrony: Tracking the gaze of viewers watching narrative virtual reality. *Media Practice and Education, 20*(3), 277–296. https://doi.org/10.1080/25741136.2018.1464743

Betz, A. (2018). The influence of the learning environment on learners' (perceived) authenticity of science communication and on their situational interest. *Unterrichtswissenschaft, 46*(3), 261–278. https://doi.org/10.1007/s42010-018-0021-0

BGHM. (2019). DGUV Regel 109–606 Branche Tischler- und Schreinerhandwerk (Issue April). Berufsgenossenschaft Holz und Metall. https://www.bghm.de/fileadmin/user_upload/Arbeitsschuetzer/Gesetze_Vorschriften/Regeln/109-606.pdf

Boucheix, J. M., Lowe, R. K., Putri, D. K., & Groff, J. (2013). Cueing animations: Dynamic signaling aids information extraction and comprehension. *Learning and Instruction, 25*, 71–84. https://doi.org/10.1016/j.learninstruc.2012.11.005

Buttussi, F., & Chittaro, L. (2018). Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE Transactions on Visualization and Computer Graphics, 24*(2), 1063–1076. https://doi.org/10.1109/TVCG.2017.2653117

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual reality sickness: A review of causes and measurements. *International Journal of Human-Computer Interaction, 36*(17), 1658–1682. https://doi.org/10.1080/10447318.2020.1778351

Clinch, S., Cortis Mack, C., Ward, G., & Steeds, M. (2021). *Technology-mediated memory impairment BT—technology-augmented perception and cognition* (pp. 71–124). Springer International Publishing.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*(2), 428–434. https://doi.org/10.1037/0022-0663.94.2.428

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Dahmani, L., & Bohbot, V. D. (2020). Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports, 10*(1), 1–14. https://doi.org/10.1038/s41598-020-62877-0

Dai, L., Jung, M. M., Postma, M., & Louwerse, M. M. (2022). A systematic review of pedagogical agent research: Similarities, differences and unexplored aspects. *Computers and Education, 190*, 104607. https://doi.org/10.1016/j.compedu.2022.104607

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1), 10–32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Davis, R. O. (2018). The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis. *Educational Research Review, 24*, 193–209. https://doi.org/10.1016/j.edurev.2018.05.002

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2011). Attention cueing in an instructional animation: The role of presentation speed. *Computers in Human Behavior, 27*(1), 41–45. https://doi.org/10.1016/j.chb.2010.05.010

Decker, D., & Merkt, M. (2023). Do a pedagogical agent's clothing and an animated video's setting affect learning? *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2023.1205338

Decker, D., & Merkt, M. (2024). Shake it or light it! The effects of cueing in desktop-VR learning environments on search time and learning. *Journal of Computer Assisted Learning, 40*(3), 1201–1217. https://doi.org/10.1111/jcal.12945

Delgado, C. Y., & Mayer, R. E. (2025). Implementing pretraining to optimise learning in immersive virtual reality. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.13099

Dolsak, W., & Reich, C. (2022). Virtual reality simulator for pipe umbrella installation works. *Geomechanik und Tunnelbau, 15*(2), 233–238. https://doi.org/10.1002/geot.202100062

Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., & Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron, 63*(1), 127–138. https://doi.org/10.1016/j.neuron.2009.06.005

Fombona-Pascual, A., Fombona, J., & Vázquez-Cano, E. (2022). VR in chemistry, a review of scientific research on advanced atomic/molecular visualization. *Chemistry Education Research and Practice, 23*(2), 300–312. https://doi.org/10.1039/d1rp00317h

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Gardony, A. L., Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2013). How navigational aids impair spatial memory: Evidence for divided attention. *Spatial Cognition and Computation, 13*(4), 319–350. https://doi.org/10.1080/13875868.2013.792821

Gardony, A. L., Brunyé, T. T., & Taylor, H. A. (2015). Navigational aids and spatial memory impairment: The role of divided attention. *Spatial Cognition and Computation, 15*(4), 246–284. https://doi.org/10.1080/13875868.2015.1059432

Gazit, N., Ben-Gal, G., & Eliashar, R. (2024). Development and validation of an objective virtual reality tool for assessing technical aptitude among potential candidates for surgical training. *BMC Medical Education, 24*(1), 1–17. https://doi.org/10.1186/s12909-024-05228-1

Grant, S. C., & Magee, L. E. (1998). Contributions of proprioception to navigation in virtual environments. *Human Factors, 40*(3), 489–497. https://doi.org/10.1518/001872098779591296

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2023). Mutual interplay between cognitive offloading and secondary task performance. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-023-02312-3

Grogorick, S., Albuquerque, G., & Magnor, M. (2018). Comparing unobtrusive gaze guiding stimuli in head-mounted displays. *Proceedings—International Conference on Image Processing, ICIP*. https://doi.org/10.1109/ICIP.2018.8451784

van der Haak, B. (2014). Digital Amnesia at Work. *Kaspersky Lab*. https://media.kasperskydaily.com/wp-content/uploads/sites/89/2016/03/05232337/Digital-Amnesia-at-work-the-risks-and-rewards-of-forgetting-in-business-report.pdf

Hazeltine, E., Teague, D., & Ivry, R. B. (2002). Simultaneous dual-task performance reveals parallel response selection after practice. *Journal of Experimental Psychology: Human Perception and Performance, 28*(3), 527–545. https://doi.org/10.1037/0096-1523.28.3.527

Hejtmánek, L., Oravcová, I., Motýl, J., Horáček, J., & Fajnerová, I. (2018). Spatial knowledge impairment after GPS guided navigation: Eye-tracking study in a virtual town. *International Journal of Human Computer Studies, 116*, 15–24. https://doi.org/10.1016/j.ijhcs.2018.04.006

Hooge, I., & Camps, G. (2013). Scan path entropy and arrow plots: Capturing scanning behavior of multiple observers. *Frontiers in Psychology, 4*, 1–10. https://doi.org/10.3389/fpsyg.2013.00996

Hornof, A. J., Zhang, Y., & Halverson, T. (2010). Knowing where and when to look in a time-critical multimodal dual task. *Conference on Human Factors in Computing Systems—Proceedings, 3*, 2103–2112. https://doi.org/10.1145/1753326.1753647

Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology, 28*(1), 74–82. https://doi.org/10.1016/j.jenvp.2007.09.002

Jackson, K. M., Shaw, T. H., & Helton, W. S. (2023). Evaluating the dual-task decrement within a simulated environment: Word recall and visual search. *Applied Ergonomics, 106*, 103861. https://doi.org/10.1016/j.apergo.2022.103861

Jackson, K. M., Shaw, T. H., & Helton, W. S. (2023b). The effects of dual-task interference on visual search and verbal memory. *Ergonomics, 66*(1), 125–135. https://doi.org/10.1080/00140139.2022.2061053

Jiang, L., Xu, M., Liu, T., Qiao, M., & Wang, Z. (2018). DeepVS: A deep learning based video saliency prediction approach. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer vision—ECCV 2018* (pp. 625–642). Springer International Publishing.

Johnson-Glenberg, M. C. (2019). The necessary nine: Design principles for embodied VR and active stem education. *Springer Singapore*. https://doi.org/10.1007/978-981-13-8265-9_5

Johnson-Glenberg, M. C., Bartolomea, H., & Kalina, E. (2021). Platform is not destiny: Embodied learning effects comparing 2D desktop to 3D virtual reality STEM experiences. *Journal of Computer Assisted Learning, 37*(5), 1263–1284. https://doi.org/10.1111/jcal.12567

Kang, Y. G., Song, H. D., Yun, H., & Jo, Y. (2022). The effect of virtual reality media characteristics on flow and learning transfer in job training: The moderating effect of presence. *Journal of Computer Assisted Learning, 38*(6), 1674–1685. https://doi.org/10.1111/jcal.12702

Khokhar, A., & Borst, C. (2022). Modifying pedagogical agent spatial guidance sequences to respond to eye-tracked student gaze in VR. Proceedings - SUI 2022: ACM Conference on Spatial User Interaction. https://doi.org/10.1145/3565970.3567697

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction. 2095–2102. https://doi.org/10.1145/2556288.2557207

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2017.01997

Landau, M., Comeaux, M., Mortell, T., Boyle, R., Imbrescia, K., & Chaffin, A. E. (2024). Characterizing the untapped potential of virtual reality in plastic and reconstructive surgical training: A systematic review on skill transferability. *JPRAS Open, 41*, 295–310. https://doi.org/10.1016/j.jpra.2024.06.015

Lange, D., Stratmann, T. C., Gruenefeld, U., & Boll, S. (2020). HiveFive: Immersion preserving attention guidance in virtual reality. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3313831.3376803

Lee, K. M. (2004). Presence, explicated. *Communication Theory, 14*(1), 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Li, W., Wang, F., & Mayer, R. E. (2023). How to guide learners' processing of multimedia lessons with pedagogical agents. *Learning and Instruction, 84*, 101729. https://doi.org/10.1016/j.learninstruc.2022.101729

Li, W., Wang, F., & Mayer, R. E. (2024). Increasing the realism of on-screen embodied instructors creates more looking but less learning. *British Journal of Educational Psychology*. https://doi.org/10.1111/bjep.12677

Li, W., Wang, F., Mayer, R. E., & Liu, H. (2019). Getting the point: Which kinds of gestures by pedagogical agents improve multimedia learning? *Journal of Educational Psychology, 111*(8), 1382–1395. https://doi.org/10.1037/edu0000352

Li, W., Wang, F., Mayer, R., & Liu, T. (2022). Animated pedagogical agents enhance learning outcomes and brain activity during learning. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12634

Lin, Y. C., Chang, Y. J., Hu, H. N., Cheng, H. T., Huang, C. W., & Sun, M. (2017). Tell me where to look. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2535–2545. https://doi.org/10.1145/3025453.3025757

Lin, L., & Atkinson, R. K. (2011). Using animations and visual cueing to support learning of scientific concepts and processes. *Computers and Education, 56*(3), 650–658. https://doi.org/10.1016/j.compedu.2010.10.007

Liu, R., Xu, X., Yang, H., Li, Z., & Huang, G. (2022). Impacts of cues on learning and attention in immersive 360-degree video: An eye-tracking study. *Frontiers in Psychology, 12*, 1–11. https://doi.org/10.3389/fpsyg.2021.792069

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology, 113*(4), 719–735. https://doi.org/10.1037/edu0000473

Makransky, G., Lilleholt, L., & Aaby, A. (2017). Development and validation of the Multimodal presence scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior, 72*, 276–285. https://doi.org/10.1016/j.chb.2017.02.066

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*(1), 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09586-2

Marsh, T., & Smith, S. P. (2001). Guiding user navigation in virtual environments using awareness of virtual off-screen space. Proceedings of the Workshop on guiding users through interactive experiences—Usability centred design and evaluation of virtual 3D environments, 149–154. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.4719

Mayer, R. E. (2014). The cambridge handbook of multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbooks in psychology.* Cambridge University Press.

Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology, 95*(4), 806–812. https://doi.org/10.1037/0022-0663.95.4.806

McKay, K. T., Grainger, S. A., Coundouris, S. P., Skorich, D. P., Phillips, L. H., & Henry, J. D. (2021). Visual attentional orienting by eye gaze: A meta-analytic review of the gaze-cueing effect. *Psychological Bulletin, 147*(12), 1269–1289. https://doi.org/10.1037/bul0000353

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part I. *Basic Mechanisms. Psychological Review, 104*(1), 3–65. https://doi.org/10.1037/0033-295X.104.1.3

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education, 140*, 103603. https://doi.org/10.1016/j.compedu.2019.103603

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition, 21*(1), 59–68. https://doi.org/10.1016/j.concog.2011.08.010

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Moritz, J., Meyerhoff, H. S., & Schwan, S. (2020). Control over spatial representation format enhances information extraction but prevents long-term learning. *Journal of Educational Psychology, 112*(1), 148–165. https://doi.org/10.1037/edu0000364

Obourdin, G., de Maeyer, S., & Van den Bossche, P. (2024). Unlocking the power of immersive learning: The FAIRI instructional design proposition for adaptive immersive virtual reality. *Computers and Education: X Reality, 5*, 100084. https://doi.org/10.1016/j.cexr.2024.100084

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*(6), 785–797. https://doi.org/10.1037/edu0000241

Parong, J., & Mayer, R. E. (2021). Cognitive and affective processes for learning science in immersive virtual reality. *Journal of Computer Assisted Learning, 37*(1), 226–241. https://doi.org/10.1111/jcal.12482

Pashler, H. (1994). Dual-task Interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220–244.

Pejsa, T., Andrist, S., Gleicher, M., & Mutlu, B. (2015). Gaze and attention management for embodied conversational agents. *ACM Transactions on Interactive Intelligent Systems, 5*(1), 1–34. https://doi.org/10.1145/2724731

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review, 17*, 19–36. https://doi.org/10.1016/j.edurev.2015.12.003

Rothe, S., spsampsps Hußmann, H. (2018). Guiding the Viewer in Cinematic Virtual Reality by Diegetic Cues. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 10850 LNCS* (pp. 101–117). https://doi.org/10.1007/978-3-319-95270-3_7

Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educational Research Review, 23*, 1–24. https://doi.org/10.1016/j.edurev.2017.11.001

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The cognitive-affective-social theory of learning in digital environments (CASTLE). *Educational Psychology Review, 34*(1), 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schüttler, T., Watzka, B., Girwidz, R., & Ertl, B. (2021). Die Wirkung der Authentizität von Lernort und Laborgeräten auf das situationale Interesse und die Relevanzwahrnehmung beim Besuch eines naturwissenschaftlichen Schülerlabors. *Zeitschrift Für Didaktik der Naturwissenschaften*. https://doi.org/10.1007/s40573-021-00128-z

Shi, L., Huang, F. C., Lopes, W., Matusik, W., & Luebke, D. (2017). Near-eye light field holographic rendering with spherical waves for wide field of view interactive 3D computer graphics. *ACM Transactions on Graphics*. https://doi.org/10.1145/31308003130832

Shu, Y., Huang, Y. Z., Chang, S. H., & Chen, M. Y. (2019). Do virtual reality head-mounted displays make a difference? A comparison of presence and self-efficacy between head-mounted displays and desktop computer-facilitated virtual environments. *Virtual Reality, 23*(4), 437–446. https://doi.org/10.1007/s10055-018-0376-x

Skulmowski, A. (2023). The cognitive architecture of digital externalization. *Educational Psychology Review, 35*(4), 1–21. https://doi.org/10.1007/s10648-023-09818-1

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers Robotics AI, 3*, 1–47. https://doi.org/10.3389/frobt.2016.00074

Smith, S. P., & Marsh, T. (2004). Evaluating design guidelines for reducing user disorientation in a desktop virtual environment. *Virtual Real., 8*(1), 55–62.

Sondermann, C., Huff, M., & Merkt, M. (2024). Distracted by a talking head? An eye tracking study on the effects of instructor presence in learning videos with animated graphic slides. *Learning and Instruction, 91*(January), 101878. https://doi.org/10.1016/j.learninstruc.2024.101878

Speicher, M., Rosenberg, C., Degraen, D., Daiber, F., & Krúger, A. (2019). Exploring visual guidance in 360-degree videos. Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video, 1–12. https://doi.org/10.1145/3317697.3323350

Stefanucci, J. K., Brickler, D., Finney, H. C., Wilson, E., Drew, T., & Creem-Regehr, S. H. (2022). Effects of simulated augmented reality cueing in a virtual navigation task. *Frontiers in Virtual Reality, 3*(September), 1–16. https://doi.org/10.3389/frvir.2022.971310

Strobach, T., Frensch, P. A., Soutschek, A., & Schubert, T. (2012). Investigation on the improvement and transfer of dual-task coordination skills. *Psychological Research Psychologische Forschung, 76*(6), 794–811. https://doi.org/10.1007/s00426-011-0381-0

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Tusher, H. M., Mallam, S., & Nazir, S. (2024). A Systematic review of virtual reality features for skill training. *Technology knowledge and learning*. Springer.

van Gog, T. (2021). The signaling (or Cueing) principle in multimedia learning. In L. Fiorella & R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 221–230). Cambridge University Press.

Volkmann, T., Wessel, D., Jochems, N., & Franke, T. (2018). German Translation of the Multimodal Presence Scale. Dachselt, R. & Weber, G. (Hrsg.), Mensch Und Computer 2018—Tagungsband, https://doi.org/10.18420/muc2018-mci-0428

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence Teleoperators and Virtual Environments, 7*(3), 225–240. https://doi.org/10.1162/105474698565686

Zhang, H., Zhang, Y., Xu, T., & Zhou, Y. (2023). Effects of VR instructional approaches and textual cues on performance, cognitive load, and learning experience. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-023-10313-1

Zhu, Y., Zhai, G., Min, X., & Zhou, J. (2020). The Prediction of saliency map for head and eye movements in 360 degree images. *IEEE Transactions on Multimedia, 22*(9), 2331–2344. https://doi.org/10.1109/TMM.2019.2957986

Daniela Decker was a PhD student at the German Institute for Adult Education – Leibniz Centre for Lifelong Learning (DIE) in Bonn, examining the instructional design of virtual-reality learning environments and educational videos. Her research at the University of Bonn currently focuses on developing an instrument to measure digital competence.

Martin Merkt is a Researcher at the German Institute for Adult Education–Leibniz Centre for Lifelong Learning in Bonn. His research focuses on the instructional design of dynamic audiovisual representations such as videos, animations, or virtual realities. He is interested in both formal and informal learning scenarios.