



How Learners' Visuospatial Ability and Different Ways of Changing the Perspective Influence Learning About Movements in Desktop and Immersive Virtual Reality Environments

Birgit Brucker¹ · Georg Pardi¹ · Fabienne Uehlin² · Laura Moosmann³ · Martin Lachmair⁴ · Marc Halfmann¹ · Peter Gerjets^{1,2}

Accepted: 20 May 2024
© The Author(s) 2024

Abstract

Virtual reality (VR) applications are developing rapidly, becoming more and more affordable, and offer various advantages for learning contexts. Dynamic visualizations are generally suitable for depicting continuous processes (e.g., different movement patterns), and particularly dynamic virtual 3D-objects can provide different perspectives on the movements. The present study investigated through a low immersive (desktop “VR”, Study 1) and a high immersive virtual environment (immersive VR; Study 2) the effectiveness of different interaction formats to view 3D-objects from different perspectives. Participants controlled either the orientation of the 3D-objects (Study 1, mouse interaction; Study 2, hand interaction via VR controllers) or their viewpoint in relation to the 3D-objects (Study 1, camera position; Study 2, position of participants' own body). Additionally, the moderating influence of learners' visuospatial ability was addressed. Dependent variables were pictorial recognition (easy, medium, difficult), factual knowledge, presence, and motion sickness. Results showed that higher-visuospatial-ability learners outperformed lower-visuospatial-ability learners. In Study 1, higher-visuospatial-ability learners showed higher recognition performance (difficult items) by controlling the camera position, whereas lower-visuospatial-ability learners suffered from this interaction format. In Study 2, higher-visuospatial-ability learners achieved better recognition performance (easy items) by controlling the 3D-models, whereas lower-visuospatial-ability learners tended to profit from moving around the 3D-objects (medium items). The immersive VR yielded more presence and higher motion sickness. This study clearly shows that different interaction formats to view 3D-objects from multiple perspectives in Desktop-VR are not transferable on a one-to-one basis into immersive VR. The results and implications for the design of virtual learning environments are discussed.

Extended author information available on the last page of the article

Keywords Immersive virtual reality · Desktop virtual reality · Dynamic 3D-models · Learners' visuospatial ability · Learning about movements

Introduction

Even though virtual reality (VR) technology and the idea to use it for learning and teaching has already existed for several decades (e.g., Bricken, 1991), VR applications and environments currently develop rapidly and, at the same time, become more and more affordable to a broader audience of possible users (e.g., Buehler & Kohne, 2019; Hilfert & König, 2016). VR applications offer various advantages not only for entertainment but also for different learning contexts (e.g., Computer Science, Pirker et al., 2020), such as experiencing impossible to attend scenarios that happened in the past (e.g., Anne Frank House in VR: Hartmann, 2013; battle of Thermopylae: Christopoulos et al., 2011), as well as scenarios that are too dangerous (e.g., virtual firefighter training: Wheeler et al., 2021), simply physically impossible to attend or experience in real life (e.g., solar system events: Huang et al., 2022; body transfer: Slater et al., 2010; virtual wings: Egeberg et al., 2016), or too time or cost expensive (e.g., virtual tourism: Wagler & Hanus, 2018; virtual scuba diving: Jain et al., 2016).

Given these advantages of VR for different scenarios in terms of interactive and engaging (learning) experiences (e.g., Hamilton et al., 2021), it has to be assumed that VR might also enhance recognition and understanding in dynamic domains that comprise different highly complex movements, such as different movement patterns of fish to generate propulsion (e.g., Imhof et al. 2012). One particularly exciting aspect of VR in this domain is its potential to offer multiple perspectives on the dynamics and movements depicted in virtual objects. Thus, the present work aims to answer whether and how interacting with different perspectives on movements affects learning in VR environments.

Learning About Movements: Dynamic Visualizations and 3D-Models

Dynamic (2D) visualizations are frequently used as instructional tools to visualize learning contents comprising continuous processes and dynamic changes because they explicitly show the visuospatial changes over time (e.g., congruence principle, Tversky et al., 2002). Dynamic visualizations prove to be highly beneficial in understanding dynamic phenomena and outperform static visualizations, particularly when learners with lower visuospatial ability must be supported (e.g., Höffler, 2010), as well as when the learning content involves biological movement (e.g., Berney & Bétrancourt, 2016; Höffler & Leutner, 2007; Plötzner et al. 2021).

Dynamic visualizations were successfully used during learning about human biological movements, such as tying knots or playing the piano, through observation (e.g., De Koning et al., 2019; Marcus et al., 2013; Mierowsky et al., 2020), but also during learning about non-human movements (e.g., fish movements, Brucker et al., 2015, Imhof et al., 2011, Imhof et al., 2012). However, dynamic visualizations

usually provide the viewers with only one perspective on the displayed (dynamic) contents or, in some rare cases, with predefined camera movements showing multiple perspectives one after another. Boucheix et al. (2018) showed that mixed camera viewpoints in instructional videos are more effective for learning complex medical hand procedures than single viewpoints or no video instructions. These findings suggest that incorporating diverse perspectives can enhance learning.

One benefit of using 3D models for educational content is that they empower learners to adjust their viewing proximity, enabling a closer and more detailed examination of specific features of the 3D model (e.g., Burkhardt et al., 2010). At the same time, these features are already ideally prepared to facilitate a comprehensive learning experience (Burkhardt et al., 2010). Moreover, interactive 3D-visualizations can encourage learners to explore the (dynamic) contents from various perspectives through interactively changing the viewpoints (e.g., interaction with animation, De Koning & Tabbers, 2011). It has been shown that 3D-models are beneficial compared to 2D-visualizations during learning (e.g., Remmele et al., 2015) and that particularly learners with higher visuospatial ability profit from 3D-models (e.g., Huk, 2006). Furthermore, an active exploration of 3D-models resulting in multiple perspectives is an effective method to promote learning (e.g., James et al., 2002; Jang et al., 2017; Keehner et al., 2008).

James et al. (2002) showed that active exploration of 3D-objects in a VR learning phase (in terms of actively rotating the 3D-objects) leads to better results in recognition performance than passively observing the objects. Jang et al. (2017) came to a similar conclusion. In their study, the learning performance of participants who directly interacted with a 3D-model of the inner ear exceeded that of those who only passively watched a video of such an exploration. According to Jang et al. (2017), interaction favors the creation of a better internal representation of the depicted virtual objects. Keehner et al. (2008) demonstrated that interactively rotating virtual objects led to better learning performance than passive observing. However, this advantage of interactivity vanished when participants were presented with the same visual input (i.e., the same multiple perspectives) in the passive observing condition.

Furthermore, in the study conducted by Keehner et al. (2008), it has been shown that participants who passively observed optimal rotations of objects, as well as those who actively interacted with the visualizations in a skillful manner, both performed better than participants who interacted with the visualizations ineffectively. Taken together, the results of Keehner et al. (2008) indicate that giving participants the possibility to interact with the visualization actively seems not to be enough. It also must be ensured that participants can perceive the most task-relevant information. Thus, the question occurs, which interaction possibilities in VR environments best allow participants to view 3D-objects from various perspectives.

Multiple Perspectives: Interaction Formats and Hand Proximity

When using 3D-models to depict dynamic virtual objects in VR, there are two different ways to change the perspective (i.e., interaction formats) on the depicted virtual objects interactively. Firstly, one could change the orientation of the virtual object

in the scene by manually rotating the object to look at it from different perspectives, whereas the position of the viewer (as well as the background of the scene) stays the same. This is called object movement in the following. Secondly, one could change the position of the viewer (e.g., by walking around the object) and thereby the view-point from where one looks at the object (as well as what is seen as a background behind the object), whereas the position of the virtual object in the scene stays the same. This is called viewer movement (or camera movement, respectively) in the following. Interacting with and moving around objects can be considered a biological primary knowledge concept (e.g., Ayres et al., 2009; Geary, 2007; Sweller, 2020; Van Gog et al., 2009). Due to evolutionary processes we—as humans—can learn and use such biological primary knowledge with minimal effort. This should hold true if the way to interact in VR environments follows the same principles that apply in the real world. However, complex interaction mechanisms, in terms of input devices that are implemented in a way that does not fit common principles, which humans are used to performing in real-life interactions, can be considered secondary knowledge (see Geary, 2007).

One factor that could potentially influence interactions with VR environments is the possible presence of the user's hands in close proximity to the processed and to-be-learned materials. Hand proximity has been linked to the idea that objects close to the hands are readily considered for immediate interaction and action (e.g., Gozli et al., 2012). The potential positive impact of hand proximity on learning might be explained through at least two principal theoretical frameworks. First, the Attentional Prioritization Theory, as proposed by Reed et al. (2006), suggests that stimuli near the hands receive enhanced visuospatial attention due to the increased activity of visuo-tactile bimodal neurons (Reed et al., 2006; see also; Brown et al., 2015; Reed et al. 2010; Tseng et al., 2012). This hypothesis is supported by further research indicating that these neurons possess overlapping tactile and visual receptive fields extending into the space surrounding the hands. Visual receptive fields near the hands exhibit a spatially graded increase in activity, with higher firing rates for objects closer to the hands, as demonstrated in studies by Graziano and Gross (1993) and Graziano et al. (1997). Graziano et al. (1994) speculated that an approximately 20-cm zone around the hands exists where bimodal neurons sensitive to both touch and visual stimuli are activated by visual input.

The second theory is the Dual Visual Pathway Theory, introduced by Gozli and colleagues in 2012. The theory explains that the proximity of objects to the hands selectively enhances the activation of one of two distinct visual processing streams identified by Goodale and Milner (1992). Specifically, it suggests that objects near the hands amplify the activity within the dorsal pathway while simultaneously diminishing the activity within the ventral pathway (Goodhew et al., 2014; Gozli et al., 2012; Taylor et al., 2015). Beyond that, Goodhew and Clarke's (2016) research discusses that the activation of magnocellular over parvocellular cells is influenced by the demands of spatial attention, suggesting that visual processing is task-dependent.

Critics have stated that no single theory comprehensively explains all observed effects of object proximity to the hands, as evidenced by various studies (e.g., Brockmole, et al., 2013; Davoli et al., 2012; Tseng & Bridgeman, 2011). The precise

processes driving the impact of hand proximity on perception remain a subject of ongoing discussion and require further exploration through primary research.

Aside from the precise reasoning for the effect, extensive research has highlighted several advantages related to hand proximity, including improved visuospatial processing, heightened attention, and enhanced visuospatial learning (e.g., Brockmole et al., 2013; Brucker et al., 2021; Reed et al., 2006; Tseng et al. 2012). In the context of using dynamic 3D-models in interactive VR environments, this leads to the question of the potential benefits of direct interactions with these models in such settings. If the mere proximity of the hands can already yield positive effects on cognitive processes and learning outcomes, then actively engaging with, and manipulating 3D-models in VR may outperform simply moving around the virtual objects without directly manipulating them with the user's hands. Understanding how direct interaction with VR elements contributes to improved comprehension and learning could have significant implications for designing and optimizing VR-based educational and training experiences. During developing more effective virtual learning environments and given the potential benefits of multiple perspectives as well as direct manipulation, it also becomes essential to explore and compare these effects in both Desktop-VR and true immersive VR environments (e.g., Johnson-Glenberg et al., 2021).

Differentiation of VR Environments

Fuchs et al. (2011) highlight the significance of understanding the purpose of virtual reality, which is to facilitate sensorimotor and cognitive activity within a digitally created artificial world. They define VR as follows: "Virtual reality is a scientific and technical domain that uses computer science (1) and behavioral interfaces (2) to simulate in a virtual world (3) the behavior of 3D entities, which interact in realtime (4) with each other and with one or more users in pseudo-natural immersion (5) via sensorimotor channels" (Fuchs et al., 2011, p. 8). Moreover, they emphasize immersion and interaction as key elements within VR environments.

However, these conditions are met by several devices and Smith (2019) differentiates between desktop and headset VR. According to Smith (2019), the term *Desktop-VR* covers all experimental setups that use an ordinary computer screen as a display where the interaction runs via a computer mouse, keyboard, or trackpad. Alternatively, alongside the term "Desktop-VR", the term virtual reality learning environments (VRLE) can also be used in this context (see Al Amiri et al., 2020; Thorsteinsson 2013).

The advantages of Desktop-VR applications lie in the low-cost procurement of the required hardware and software as well as the familiarity of the participants with these input devices (thereby avoiding long familiarization and practice phases; Smith, 2019). In Desktop-VR settings, content is displayed on a two-dimensional screen, limiting depth perception to monocular (instead of stereoscopic) cues. Furthermore, the use of mouse and keyboard for interaction significantly differs from the real-world motor activities being simulated (Smith, 2019).

In contrast, according to Smith (2019), in *headset VR* environments (called *immersive VR* in the following), the digitally generated environment is projected directly onto a head-mounted display (HMD), which users wear directly in front of their eyes, and the interaction runs in these environments via controllers or hand-based gesture interactions (e.g., Li et al., 2019) represented in the VR as virtual hands. HMDs not only allow for stereoscopic vision by presenting the two eyes with images differing minimally in perspective (as is the case with natural vision due to human anatomy) but also interpret the (head) movements of the user and adapt the visual information accordingly (Smith, 2019).

Immersion and Presence in VR Environments

Independent of the differentiation of VR systems, two central characteristics of VR environments are *immersion* and *presence* (see Slater & Wilbur, 1997). *Immersion* is defined by Slater and Wilbur (1997) as an objectively detectable technological property with the following subcategories: (1) inclusiveness (degree to which the physical reality is excluded), (2) extensiveness (number of sensory modalities represented), (3) surroundedness (spatial extent of the VR), (4) vividness (degree of resolution, fidelity, and quality of the stimuli presented), (5) matching (degree of correspondence between proprioceptive information received by users and information provided by the display), and (6) plot (extent of a plot in VR that is self-contained and clearly distinguishable from reality, including autonomy as the degree of independent behavior of objects and interaction as the degree to which users can modify what is happening).

Recently, the emphasis on the concept of *presence* has become more pronounced with the advent of immersive technologies (see Cummings and Bailenson 2016). The growing emphasis on presence is attributable to the necessity of comprehending the psychological effects of recent technological innovations and the superior experiences they can produce. According to Slater and Wilbur (1997), presence describes the subjectively perceived feeling of “being there”. Current research is diverse regarding the conceptual description of presence (Grassini & Laumann, 2020). According to Hartmann and colleagues (2015), two pivotal elements of presence are the social and spatial dimensions. In the present context of one person learning within VR without social interaction, the spatial presence is of particular interest: Spatial presence is described as the subjective experience of a user perceiving themselves to be physically placed “within” a digitally mediated space, notwithstanding its origin as a technology-crafted illusion (Hartmann et al., 2015). Schubert et al. (2001) assume that (spatial) presence arises through constructing a spatial-functional-mental model of the non-physical environment. The cognitive processes involved are, on the one hand, the formation of mental representations of one’s own body movements in virtual space and, on the other hand, the suppression of information from physical space that is contrary to the impressions from the virtual environment (Grassini & Laumann, 2020; Schubert et al., 2001). Slater and Wilbur (1997) describe presence as an increasing function of immersion. Thus, as the degree of immersion increases, users will likely feel more present in the situation.

However, immersive VR environments might be perceptually more demanding than Desktop-VR environments (see Skulmowski, 2023a). Overall, it can be stated that Desktop-VR environments provide a lower level of immersion and, in line with that, a lower level of presence than immersive (headset) VR environments (Smith, 2019). In line with that, Johnson-Glenberg et al. (2023) showed in one study that more immersive VR conditions (large 2D- and 3D-screens) outperform a small 2D-screen-condition on learning outcomes. However, not only the degree of immersion and level of presence are influencing factors on the effectiveness of different VR environments, but also the embodiment and the resulting agency might play a role (as shown by another study with a more detailed result pattern by Johnson-Glenberg, 2018).

Embodiment and Agency in VR Environments

Johnson-Glenberg (2018) further emphasizes a third fundamental aspect of immersive VR environments: embodiment and the resulting *agency* linked to content manipulation in three dimensions. This aspect pertains to the level of personal empowerment or control that users have over the digital content. Embodiment and agency are fostered by developing new VR hand controllers enabling meaningful and congruent movements that align with the to-be-learned contents (Johnson-Glenberg, 2018). Skulmowski and Rey (2018) highlight in their taxonomy on embodied learning the importance of both bodily engagements, in terms of to what extent bodily activity is involved during learning, as well as task integration, in terms of whether bodily activities are meaningfully related to the respective learning tasks or not. Lachmair et al. (2022) emphasize that the whole body of a person can potentially be involved in immersive VR interactions and that this resulting amount of interaction possibilities might overwhelm users.

Makransky and Petersen (2021) introduce the Cognitive Affective Model of Immersive Learning (CAMIL) that also highlights presence and agency as the key benefits of immersive VR, because these aspects are facilitated by immersion, interactive control, and representational fidelity (see also Dalgarno & Lee, 2010). CAMIL identifies six factors influencing learning that can be enhanced by presence and agency: interest, motivation, self-efficacy, embodiment, cognitive load, and self-regulation (Makransky & Petersen, 2021, see also Petersen et al., 2022).

Fischer and Brugger (2011) present a model for understanding cognition, particularly emphasizing the role of the body by introducing the GES—an acronym for grounding, embodiment, and situatedness—framework. The GES framework has also been recently used as a cognitive framework that can help to inform interaction guidelines for user interface design, especially in VR (see Lachmair et al., 2022). Grounding refers to universal principles applicable to all individuals, such as physical laws (e.g., gravity, object impermeability, causality), which constitute a foundational element of our cognitive architecture and are resistant to change. Embodiment captures the unique, individual-specific characteristics, including bodily traits (e.g., handedness, sensory-motor proficiency), highlighting the variability in cognitive experiences based on physical constitution. Lastly, situatedness encompasses the

context-specific experiences arising from interaction within a particular environment, underscoring how situational factors profoundly shape cognitive outcomes (see Fischer & Brugger, 2011).

Research conducted thus far has yielded inconclusive results when comparing Desktop-VR and immersive VR technologies. A recent meta-analysis demonstrated beneficial effects for Desktop-VR (e.g., Cromley et al., 2023), whereas another meta-analysis indicated advantages for immersive VR (e.g., Wu et al., 2020). Johnson-Glenberg et al. (2021) investigated whether learning in a 3D VR environment is superior to learning through a 2D monitor of a personal computer (PC): participants on both platforms (3D VR versus 2D PC) learned either by watching a playback video (low level of embodiment and agency) or by using controller or mouse to manipulate the content (high level of embodiment and agency). Results revealed that higher embodiment and agency in terms of active manipulation were beneficial. However, immersive VR (3D VR) did not simply outperform Desktop-VR (2D PC) platforms because the low embodied VR group performed significantly worse than the three other groups. In contrast, the high embodied VR group demonstrated the highest level of learning and retention (Johnson-Glenberg et al., 2021). This discrepancy may help elucidate the inconclusive findings obtained thus far. Nonetheless, investigating higher levels of embodiment and agency in VR needs to be further explored.

There are various ways of interactive control, that enable embodiment and agency in different ways (i.e., object movement versus viewer movement), and they need to be examined to identify differences between them as well as differences across alternative platforms, as it cannot be assumed that results from one platform can be directly applied to the other. The exploration of interactive control with perspectives in different VR environments constitutes a major objective of this paper. Through this investigation, we address how the immersive nature of VR environments influences users' ability to engage with and manipulate perspectives within the VR environment, with particular emphasis on its implications for learning.

Beyond the ability of interaction and the differentiation of VR, as we will outline in the following, visuospatial ability plays a pivotal role in learning about dynamic phenomena.

Learners' Visuospatial Ability and Learning About Movements

When considering learning about movements with dynamic visualizations, 3D-models, and VR environments, learners' visuospatial ability emerges as a crucial factor to be considered. Understanding how learners' characteristics in terms of their visuospatial ability interact with different interaction formats and how they influence the learning process within different VR applications can provide valuable insights for designing learning environments. As Mayer and Sims (1994) described, visuospatial ability encompasses mental manipulation and visualization of spatial information about objects in two or three dimensions. This includes mentally rotating objects, understanding spatial relationships, and interpreting 3D-representations (e.g., Hegarty & Waller, 2005). These abilities are crucial for observing continuous

dynamic changes of 3D-models from different perspectives in VR environments (e.g., Sun et al., 2019). Extensive research on learners' visuospatial ability (e.g., Höffler, 2010) consistently demonstrates that learners with higher visuospatial ability outperform those with lower visuospatial ability when engaged in learning with visualizations. This suggests that visuospatial abilities are critical in effectively comprehending and utilizing visual learning materials. Moreover, previous research reveals that visuospatial ability may moderate learning effectiveness with different instructional formats (e.g., dynamic versus static visualizations, e.g., Höffler, 2010; 2D- versus 3D-models, e.g., Huk, 2006; or different VR environments, e.g., Sun et al., 2019). There are two alternative interaction assumptions: the *ability-as-compensator* versus the *ability-as-enhancer* assumption (e.g., Höffler, 2010).

Regarding the *ability-as-compensator* assumption, learners with higher visuospatial ability may not require well-designed instructional formats as they achieve high learning outcomes even when exposed to suboptimal instructional formats. On the other hand, learners with lower visuospatial ability may struggle when faced with suboptimal designed instructions (see *ability-as-compensator* assumption; Höffler, 2010). As a result, specific instructional formats (e.g., dynamic visualizations) act as compensatory tools for learners with lower visuospatial ability, enabling them to achieve similar learning outcomes to learners with higher visuospatial ability (e.g., Lee, 2007). In contrast, the *ability-as-enhancer* assumption suggests that learners with higher visuospatial ability might even be able to utilize specific instructional formats, even if these formats are less optimally designed (for example due to their necessary complexity). Their enhanced visuospatial capabilities facilitate a more effective engagement with such specific instructional formats, which could result in improved learning outcomes. For example, their advanced imagination may enable a more accurate evaluation of how a model can be rotated to achieve an optimal viewing position. In sum, considering learners' visuospatial ability is relevant when studying different interaction formats in different VR environments during learning about movements.

Present Research

The present research investigated the effectiveness of different interaction formats with dynamic 3D-models, allowing observation from multiple perspectives in desktop and immersive VR settings. To achieve these objectives, we addressed two different interaction formats (object movement versus camera/viewer movement) in two separate studies. Study 1 focused on a Desktop-VR environment, whereas Study 2 focused on an immersive VR environment. Additionally, in Study 1, we included a *no interaction* condition, in which the 3D models were displayed dynamically on the screen but without participants having the opportunity to interact with them by influencing neither the environment nor the object. Beyond the effect of the interaction formats, the potential moderating influence of learners' visuospatial ability during learning through different interaction formats was explored.

With regard to the interaction format *object movement*, the possibility to directly manipulate and rotate the virtual 3D-models allowed to change the

viewpoint and thereby observe different perspectives on the depicted dynamic content. At the same time, the viewer stayed in the same position. In contrast, during the interaction format *camera/viewer movement* the possibility to move the camera or the body of the viewer around the virtual 3D-model allowed to change the viewpoint and the resulting different perspectives. At the same time, the position and orientation of the 3D-model in the virtual scene stayed the same.

Within Study 1, both interaction formats were controlled via the mouse as input device. In contrast, in Study 2, the interaction formats were controlled in the object movement condition via VR controllers and in the viewer movement condition via body movements by walking around the 3D-model in the virtual scene. We have formulated the following hypotheses.

Hypothesis 1 (H1): object movement > camera/viewer movement > no interaction

We were interested in the effectiveness of the two interactive formats in a desktop and a truly immersive VR environment. We expected that the possibility to directly manipulate the orientation of the virtual 3D-objects (object movement) would outperform the possibility of changing the perspective from which one is looking at the 3D-models by changing the viewpoint of the camera or the viewer (camera/viewer movement) due to possible effects of hand proximity. This is expected due to incorporating the dynamic 3D-models into the peripersonal space around the hands, thereby possibly making available additional cognitive resources for learning and, thus, resulting in a more accurate dynamic mental model. Moreover, we expected that both interactive formats outperform a condition where no interaction is allowed (i.e., the dynamics must be observed from a given predefined perspective).

Hypothesis 2 (H2): higher visuospatial ability > lower visuospatial ability

We expected a main effect for learners' visuospatial ability: learners with higher visuospatial ability will be better able to mentally manipulate and understand the 3D-models and virtual scenes than learners with lower visuospatial ability. This should improve learning outcomes for learners with higher visuospatial ability.

Research Question 1 (R1): moderating role of visuospatial ability – enhancer or compensator

Additionally, we explored how learners' visuospatial ability might moderate the effectiveness of various interaction formats, examining whether these abilities act more as compensators or enhancers according to the respective assumptions.

Finally, we were interested in the question whether these effects are the same in different VR environments by exploring these in a Desktop-VR in Study 1 and an immersive VR in Study 2.

Study 1

Method

Participants and Design

In total, 164 participants were recruited through an online platform (www.prolific.com; selection criteria: native German speaker; age 18–40 years) from which they were redirected to IWM-study (Klemke & Halfmann, 2020), a specially developed online test environment accessible through a browser (compatible with all commonly used web browsers). The questionnaires and the unity-based Desktop-VR learning environment were presented to the participants there. Due to technical issues, four individuals had to be excluded from the analysis, resulting in a final sample size of 160 participants (99 males, 57 females, 4 diverse; 140 right-handed, 18 left-handed, 2 ambidextrous; 74 university students, 68 employed professionals, 17 identified as either job seekers, unemployed, pupils, or housewives/husbands, 1 missing for occupation) with an average age of 27.26 years ($SD = 5.53$). Participants were reimbursed £8.75 for approximately 45 min. The study was approved by the local ethics committee of the Leibniz-Institut für Wissensmedien (IWM, LEK 2020/024). Participants were randomly assigned to one of three conditions in a between-subject design investigating the influence of the first independent factor *interaction format* with the three factor levels: first, camera movement ($n = 54$), second, object movement ($n = 53$), and third, no interaction condition ($n = 53$; see a detailed description of the three different conditions below in the section on "interaction formats"). Moreover, participants' *visuospatial ability* was assessed as a second continuous independent factor (higher visuospatial ability was operationalized as one standard deviation above the mean, whereas lower visuospatial ability was operationalized as one standard deviation below the mean). As dependent variables, we measured pictorial recognition (easy, medium, difficult), factual knowledge about movement patterns (see Appendix 1 for the factual knowledge questions), feeling of presence, and motion sickness.

Materials and Domain

Within the Desktop-VR learning environment, participants were exposed to highly realistic 3D-models of four fish, visualizing the movement patterns these fish use for propulsion. Identifying these movement patterns in real fish is challenging, given that fish may also deploy other movements (e.g., for navigational purposes). This complex dynamic learning domain has already been investigated and tested in previous studies (e.g., Imhof et al., 2012), in which dynamic 2D-visualizations have been used that were rendered based on the 3D-models used in the present study. Participants were asked to study the 3D-models to learn how to classify fish based on the used body parts (i.e., several fins or the body itself) and how they move in the three-dimensional space (i.e., paddle-like or wave-like). Four different movement patterns were shown (i.e., *subcarangiform*, *balistiform*, *labriform*, and *tetraodontiform*). In

this study, the 3D-models were integrated into a virtual underwater environment containing no elements except a sandy bottom and the surrounding water. In addition to the 3D-models, verbal explanations of the movement patterns were provided. These spoken texts were consistent across all experimental conditions and covered the principle of locomotion (undulation or oscillation), physical parameters of the movement (amplitude and wavelength), body parts involved, and the typical fish species associated with each movement pattern. The verbal information was not synchronized with the movements of the 3D-models, as it referred to more general characteristics of the movement patterns and was not time-dependent.

Interaction Formats

The three different interaction formats, *object movement*, *camera movement*, and *no interaction*, were implemented as follows. In the *object movement condition* (see Fig. 1), participants were given the ability to manipulate the orientation of the 3D-model in the virtual learning environment by rotating it via mouse interaction (movement of the virtual object, no movement of the camera/viewer).

To modify the orientation, users needed to position the mouse cursor over the 3D-model and then press and hold the left mouse button to “lock” the cursor in place. Releasing the mouse button preserved the current orientation of the 3D-model, ensuring it was retained in the last adopted position. Because the 3D-models needed to be rotatable on three axes, but the mouse as an input device operates only in a two-dimensional space, we implemented the following types of rotation to enable participants to rotate the model freely on all three dimensions (Fig. 2).

1. Rotation around the pitch axis: By holding down the right mouse button and circularly moving the mouse, the 3D-model rotated on the vertical axis like being placed on a steering wheel.
2. Rotation around the roll axis: By holding down the left mouse button and moving the mouse up and/or down, the 3D-model rotated away from and/or towards the viewer. This action figuratively rolled the 3D-model.
3. Rotation around the yaw axis: By holding down the left mouse button and moving the mouse to the left and/or right, the 3D-model rotated like being placed on a turntable.

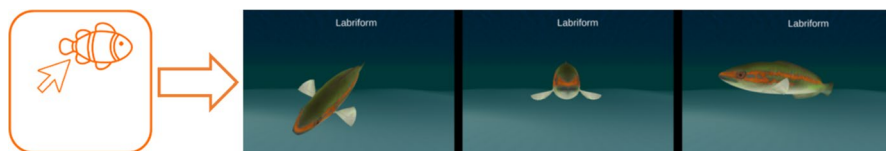


Fig. 1 These static screenshots of different perspectives in the object movement condition were obtained through the manipulation of the object’s position and orientation. The background remained constant as a point of reference, because only the object is rotated, while the camera position stayed stable.

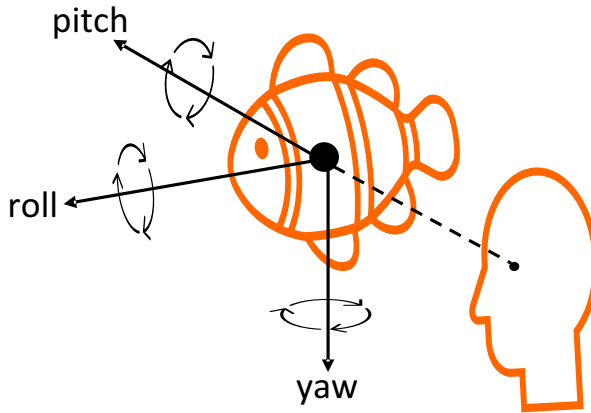


Fig. 2 Representation of the pitch, roll, and yaw axis on which the model could be rotated. The initial position of the observer relative to the object was lateral.

In the *camera movement* condition (see Fig. 3), participants could manipulate the camera's perspective on the 3D-model within the virtual learning environment via mouse interaction (this time: movement of the camera/viewer, no movement of the virtual object).

To adjust the viewpoint of the camera and thereby also the perspective from which the viewer looked at the virtual object, the mouse cursor had to be positioned anywhere in the virtual scene and be “locked” to it by pressing and holding the left mouse button while simultaneously moving the mouse (up and down as well as left and right) to change the camera's position. Releasing the left mouse button would maintain the current perspective. The camera's distance was held constant on an invisible sphere surrounding the 3D-model in its center. Moreover, the camera's orientation was always towards the fish. Furthermore, the camera and thereby the view on the whole scene could not be rotated on the pitch axis, so the bottom could never become sloped or even upside down. Therefore, no implementation of the third rotation axis (pitch axis) was needed in this condition. In the no interaction condition without interactivity, the perspective on the 3D-models corresponded to the perspective used in previous studies (e.g., Imhof et al., 2012), and participants were not able to interact with the 3D-models in the Desktop-VR environment.

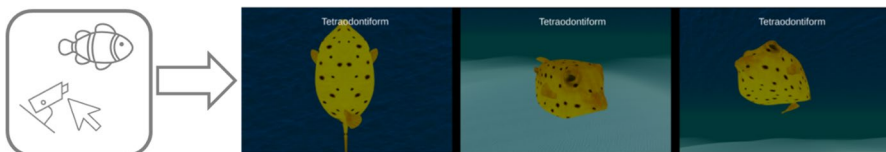


Fig. 3 These static screenshots of different perspectives in the camera movement condition were obtained through the manipulation of the camera's position and orientation. The background changed as a point of reference, because the camera position changed, while the fish position in the virtual scene was fixated.

Measures

Dependent Variables for Learning Outcomes: Recognition and Factual Knowledge First, we administered a movement pattern recognition test comprising 45 dynamic multiple-choice items to assess recognition as a learning outcome. These items (see Fig. 4) consisted of underwater videos of real fish in their natural environment performing one of the four to-be-learned or distractor movement patterns. The videos were displayed system-paced and lasted 7 s each. After 7 s, the video disappeared, leaving only the five answer options visible (one for each of the four to-be-learned movement patterns and the additional option “none of the above”). To correctly identify the movement pattern, learners had to identify the used body parts relevant for propulsion as well as how they were moved.

The videos were categorized into three difficulty levels: easy, medium, or difficult. The selection of test videos followed the methodology outlined in the study by Imhof et al. (2012), which categorized the videos based on factors such as visibility of the movement pattern, absence of secondary movements (e.g., for navigation), and similarity to other movement patterns. Easy videos showed the relevant movement pattern solely and continuously (6 videos). Medium videos continuously showed the relevant movement, but additional navigational movements occurred (20 videos). Difficult videos did not continuously show the relevant movement pattern and included several additional movements (19 videos). The score for each difficulty level was calculated by summing up the number of correct answers in the category. Each correct answer was awarded one point, and the percentage correct was calculated.

Secondly, the study assessed participants’ factual knowledge about locomotion principles employed in the movement patterns (undulation or oscillation), the body parts involved, and the physical parameters of the movement (such as amplitude). This knowledge was evaluated through eight multiple-choice questions (see Appendix 1). The necessary information for answering these questions was provided in the verbal materials of the study (consistent across all experimental conditions).

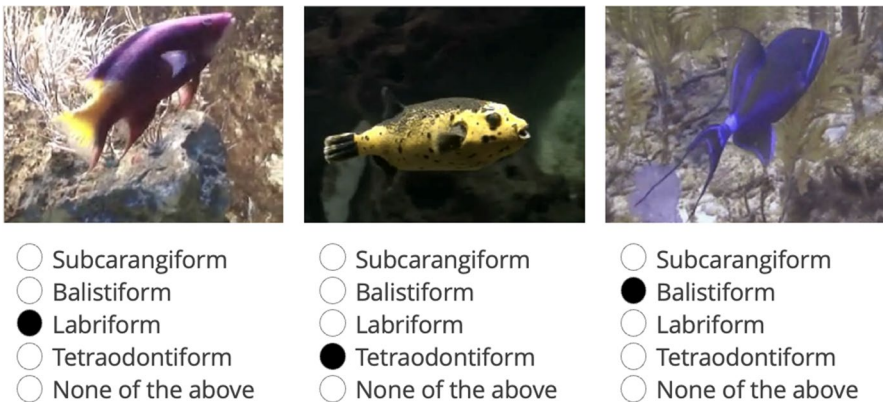


Fig. 4 Example screenshots of easy, medium, and difficult recognition items (from left to right)

Participants were informed that multiple correct answers could exist without specifying the exact number. Participants were asked two questions for each movement pattern, each offering four possible answer options, of which two were correct.

Assessment of Learners' Visuospatial Ability We used a short version of the paper folding test to assess the second independent factor—learners' visuospatial ability (PFT, Ekstrom et al., 1976). Following Blazhenkova and Kozhevnikov (2009, p. 640), this test measures the ability to form representations of “object location, movement, spatial relationships, and transformations”. This makes the PFT well-suited to cover the domain of fish movements. The short version of the PFT comprises ten multiple-choice items, where participants view depictions of papers being folded and punched. Participants must choose the correct answer from five options showing unfolded papers with punches in various positions. Correct answers earn one point (max. 10 points). Participants have 3 min to complete the task.

Additional Variables: Familiarity with the Domain, Motion Sickness, and Presence In line with prior studies (e.g., Brucker et al., 2015) in this domain, we assessed, besides demographical data (age, gender, handedness, occupation) also, participants' familiarity with the domain (as an indicator of prior knowledge that might affect learning outcomes in VR environments, e.g., Dengel & Mägdefrau, 2021). The questions for familiarity with the domain asked about participants' biology school education, experiences with marine activities like diving, snorkeling, swimming, rowing, or owning an aquarium, interest in related topics such as fish, biology, zoology, physics, aircraft construction, and shipbuilding, as well as their exposure to related media like documentaries, books, or aquarium visits. The questionnaire comprised 22 questions and was developed with input from a domain expert. Participants received one point for each indication of familiarity with the domain and could earn additional points for higher numbers of diving and/or snorkeling experiences (max. 32 points).

Simulator motion sickness caused by the VR environment was measured with the Virtual Reality Sickness Questionnaire (VRSQ) by Kim et al. (2018). The VRSQ comprises nine items all rated on a five-point Likert scale. Participants' perceived presence was measured with the Igroup Presence Questionnaire (IPQ) developed by Schubert et al. (2001). The IPQ consists of 14 questions in total: one question captures the general presence, whereas the other assesses three specific components of presence: *spatial presence*, *involvement*, and *realness* of the environment (on seven-point Likert scales specific to each question).

Procedure Participants were recruited through the online platform Prolific (London, UK). They were required to use a computer or laptop (touch devices were excluded) equipped with active loudspeakers and an external mouse (a touchpad was not permitted). Participants who confirmed these criteria were directed from Prolific to IWM-study. The study comprised a preliminary phase, a learning phase in the Desktop-VR environment (where the experimental manipulation took place),

the assessment of additional variables, and a testing phase. Participants navigated through the study using buttons. In the preliminary phase, participants were provided with general information about the study and were required to confirm their participation eligibility and informed consent. Following the initial instructions, participants were instructed to proceed with the study in full-screen mode for optimal viewing. Before starting the study itself, a VR check of the Unity application was conducted to ensure that the hardware being used met the requirements of the Unity application. If the hardware did not pass the VR check, the participants could not proceed with the Study and were redirected to Prolific. If the VR check was passed, a questionnaire was administered (demographics, familiarity with the domain). Additionally, visuospatial ability was measured with the PFT (Ekstrom et al., 1976), from which participants were automatically directed to the next page after 3 min.

Following the preliminary phase, participants were given instructions on how to proceed during the learning phase. They were informed about the respective interaction format in the Desktop-VR depending on their experimental condition. The learning phase started with an exercise to familiarize participants with the respective interaction format. For this purpose, participants were provided with a 3D-model of a submarine. In both interactive conditions (object movement and camera movement), participants were allowed to practice their respective way of interacting (object movement: manipulating the orientation of the 3D-model versus camera movement: changing the camera's viewpoint) while receiving a 72-s verbal explanation about submarines. This practicing task lasted at least two and at the maximum 5 min. After 2 min, participants could choose to proceed with the study by clicking the respective next button. Participants in the no interaction condition saw the submarine without interaction possibilities for 2 min and heard the verbal explanation.

Subsequently, participants in all experimental conditions were provided with a written introduction to "underwater locomotion", covering the basic principles of locomotion (undulation, oscillation), the body parts possibly involved in different movements, and some physical parameters of fish movements (amplitude, wavelength). Then, the four 3D-models of the to-be-learned movement patterns were presented one after another each for 2 min, accompanied by a 72-s verbal explanation that provided information about the specific movement pattern, body parts involved, and its physical parameters. Participants could pause briefly between each movement pattern before clicking the next button to see the following pattern. Depending on their respective experimental interaction format conditions, participants could or could not interact with the 3D-models during the presentation of the movement patterns. Following the learning phase, participants had to answer the VRSQ (Kim et al., 2018) and the IPQ (Schubert et al., 2001). In the following testing phase, learning outcomes in terms of recognition and factual knowledge were assessed. Finally, participants were informed about the study goals and redirected to Prolific.

Results

All data (in both studies) were analyzed with the statistical analysis software SPSS® (version 28, IBM Corp. released in 2022). A *p*-value of .005 was used to indicate

significance and a p -value of .100 was used to indicate a (non-significant) tendency in all analyses (in both studies). Regarding the comparability between the experimental conditions, we found no significant differences (all $ps > .282$) between groups (see Appendix Table A1 for means and values) in participants' age, gender, handedness, occupation, familiarity with the domain, or visuospatial ability.

Learning Outcomes

Recognition performance in terms of easy, medium, and difficult recognition, as well as factual knowledge were analyzed with ANCOVAs with the factor interaction format (object movement versus camera movement versus no interaction) and the second continuous factor learners' visuospatial ability. The continuous factor learners' visuospatial ability was z-standardized. The interaction term between interaction format and learners' visuospatial ability was inserted into the model of all analyses to test the moderating effect of learners' visuospatial abilities on the different interaction formats. Higher visuospatial ability was defined as the mean plus one standard deviation, whereas lower visuospatial ability was defined as the mean minus one standard deviation. Moreover, learners' familiarity with the domain was added as a covariate (without interaction terms) into the model of all analyses (see Table 1 for adjusted means and standard errors).

There was no significant main effect of interaction format on easy recognition, medium recognition, difficult recognition (all $F_s < 1$), and factual knowledge ($F(2, 153) = 1.128$, $MSE = 139.033$, $p = .326$, ns, $\eta^2_p = .002$). Contrary to our hypothesis H1, we observed no overall positive effect of interactivity or direct manipulation in the object movement condition compared to the camera movement condition on learning outcomes.

Regarding learners' visuospatial ability, there was a significant main effect on all three difficulty levels of recognition performance (easy: $F(1, 153) = 4.122$, $MSE = 740.494$, $p = .044$, $\eta^2_p = .026$; medium: $F(1, 153) = 4.299$, $MSE = 370.965$, $p = .040$, $\eta^2_p = .027$; difficult: $F(1, 153) = 12.111$, $MSE = 243.692$, $p < .001$, $\eta^2_p = .073$) as well as on factual knowledge ($F(1, 153) = 9.182$, $MSE = 139.033$, $p = .003$, $\eta^2_p = .057$). Participants with higher visuospatial ability achieved higher recognition performance and factual knowledge than those learners with lower visuospatial ability. This finding supports our hypothesis H2, suggesting that higher visuospatial ability positively influences learning performance.

Moreover, there was a significant interaction between interaction format and learners' visuospatial ability for difficult recognition performance ($F(2, 153) = 3.328$, $MSE = 243.692$, $p = .038$, $\eta^2_p = .042$). Bonferroni-adjusted pairwise comparisons revealed that even though the overall interaction pattern reached statistical significance ($p = .038$; see Fig. 5), the pairwise post-hoc comparisons were not statistically significant (all $ps > .102$). Additionally conducted LSD-adjusted pairwise comparisons revealed that participants with higher visuospatial ability achieved higher values for difficult recognition when using the camera movement condition compared to the no interaction condition ($p = .034$), but not compared to the object movement condition ($p = .120$), whereas there were no differences for participants with lower visuospatial ability for difficult recognition (comparison camera movement and no

Table 1 Means and standard errors (in parentheses) for easy, medium, and difficult recognition and factual knowledge (in % correct) for higher and lower visuospatial ability (VSA) in all three experimental conditions of Study 1

	Object movement ($n = 53$)				Camera movement ($n = 54$)				No interaction ($n = 53$)			
	Higher VSA		Lower VSA		Higher VSA		Lower VSA		Higher VSA		Lower VSA	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Easy	72.33	5.61	61.96	5.31	72.66	5.41	66.41	5.42	72.67	4.96	62.70	5.18
Medium	70.72	3.97	65.95	3.76	72.56	3.83	68.67	3.83	72.94	3.51	65.39	3.67
Difficult	59.53	3.22	54.73	3.05	66.54	3.10	48.62	3.11	57.55	2.84	54.12	2.97
Factual knowledge	71.24	2.32	66.97	2.30	75.13	2.35	67.55	2.35	70.65	2.15	65.30	2.25

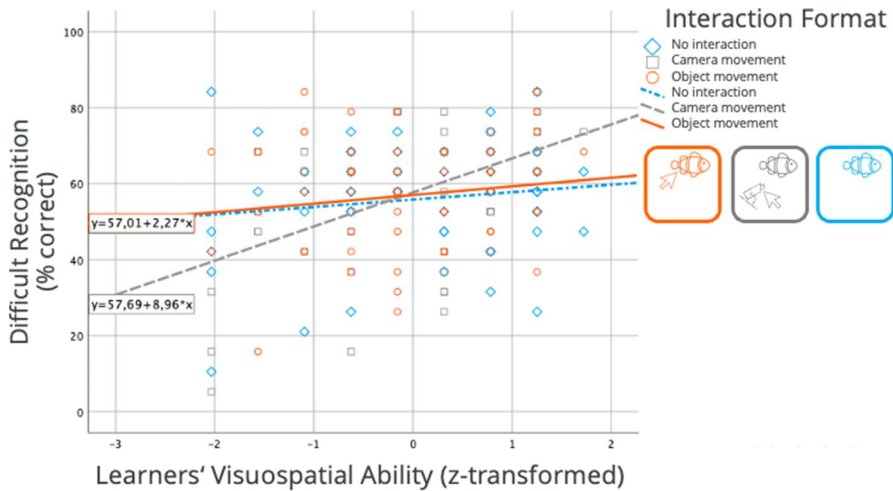


Fig. 5 Interaction between interaction format and learners' visuospatial ability on difficult recognition in Study 1

interaction condition, $p = .204$; comparison camera movement and object movement, $p = .162$). There were no other significant main effects or interactions concerning learning outcome measures (all $ps > .317$).

Motion Sickness and Presence

Motion sickness as well as the four subcategories of presence (general presence item, spatial presence, involvement, and realness) were also analyzed with ANCOVAs with the factor interaction format (object movement versus camera movement versus no interaction) and the second continuous factor learners' visuospatial ability (z-standardized with interaction term) as well as learners' familiarity with the domain as a covariate (without interaction terms; see Table 2 for adjusted means and standard errors).

There was no main effect for interaction format ($F(2, 153) = 1.151$, $MSE = .289$, $p = .319$, ns, $\eta_p^2 = .015$) on motion sickness. However, there was a main effect of learners' visuospatial ability on motion sickness ($F(1, 153) = 5.055$, $MSE = .289$, $p = .026$, $\eta_p^2 = .032$) in the direction that learners with higher visuospatial ability experienced lower motion sickness than learners with lower visuospatial ability. Moreover, in addition to this main effect of visuospatial ability there was a (non-significant) tendency for an interaction between interaction format and learners' visuospatial ability for motion sickness ($F(2, 153) = 2.850$, $MSE = .289$, $p = .061$, ns, $\eta_p^2 = .036$; see Fig. 6). Bonferroni-adjusted pairwise comparisons revealed that for learners with lower visuospatial ability the no interaction condition led to higher motion sickness than the camera movement condition ($p = .036$), but not than the object movement condition ($p = .251$); whereas there were no differences for learners with higher visuospatial ability (all $ps > .999$).

Table 2 Means and standard errors for motion sickness, general presence item, spatial presence, involvement, and realism for higher and lower visuospatial ability (VSA) in all three experimental conditions of Study 1

	Object movement (<i>n</i> = 53)				Camera movement (<i>n</i> = 54)				No interaction (<i>n</i> = 53)			
	Higher VSA		Lower VSA		Higher VSA		Lower VSA		Higher VSA		Lower VSA	
	M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Motion sickness (1-5)	1.44	0.11	1.65	0.11	1.58	0.11	1.52	0.11	1.47	0.10	1.90	0.10
General presence (1-7)	3.03	0.29	3.17	0.28	3.22	0.28	2.62	0.25	2.70	0.26	2.26	0.27
Spatial presence (1-7)	3.26	0.22	3.30	0.21	3.68	0.21	3.52	0.21	3.14	0.20	2.80	0.20
Involvement (1-7)	3.37	0.25	3.41	0.24	3.20	0.25	3.06	0.25	3.01	0.22	2.82	0.23
Realness (1-7)	2.34	0.20	2.64	0.19	2.50	0.19	2.47	0.19	2.54	0.17	2.18	0.18

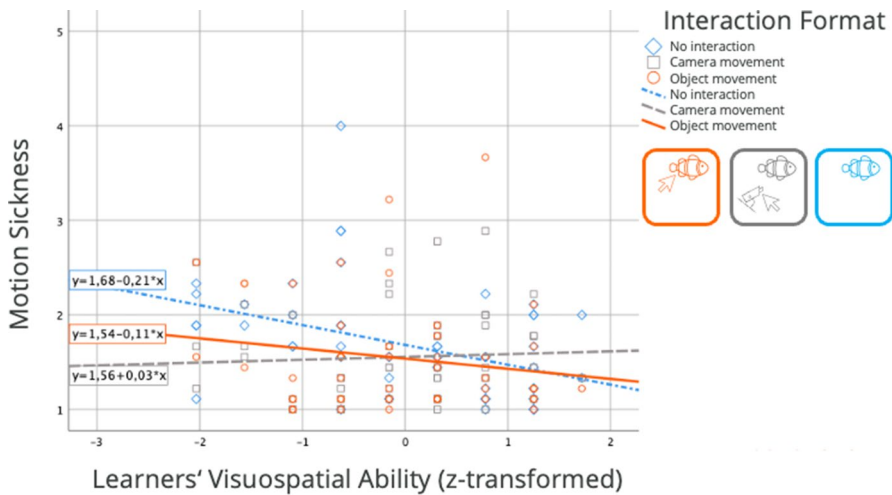


Fig. 6 (Non-significant) tendency for an interaction between interaction format and learners' visuospatial ability on motion sickness in Study 1

There was a significant main effect of interaction format on spatial presence ($F(1, 153) = 4.604$, $MSE = 1.150$, $p = .011$, $\eta^2_p = .057$). Bonferroni-adjusted pairwise comparisons showed that camera movement led to higher spatial presence ratings than the no interaction condition ($p = .009$), whereas there was no difference between object movement and no interaction ($p = .427$). Moreover, there was a (non-significant) tendency for the general presence item ($F(1, 153) = 2.700$, $MSE = 1.989$, $p = .070$, ns, $\eta^2_p = .034$). Bonferroni-adjusted pairwise comparisons showed that object movement tended to lead to higher ratings for the general presence item than the no interaction condition ($p = .077$), whereas there was no difference between camera movement and no interaction ($p = .325$).

Moreover, there was a significant main effect of familiarity with the domain on the general presence item ($F(1, 153) = 9.361$, $MSE = 1.989$, $p = .003$, $\eta^2_p = .058$) and spatial presence ($F(1, 153) = 4.884$, $MSE = 1.150$, $p = .029$, $\eta^2_p = .031$). Higher familiarity with the domain leads to higher presence ratings. There were no other main effects or interactions concerning the four subcategories of presence (all $ps > .266$).

Discussion of Study 1

In the first study, we set out to investigate the influence of different interaction formats (object movement, camera movement, no interaction) on learning performance in a Desktop-VR setting. Our initial hypothesis (H1) suggested that object movement would lead to better learning outcomes compared to camera movement, and both interactive formats would outperform the no interaction condition. Additionally, we hypothesized in hypothesis H2 that learners with higher visuospatial ability would outperform those with lower visuospatial ability across all learning outcome measures (easy, medium, and difficult recognition, and factual knowledge).

However, contrary to our hypothesis H1 on interaction format, we did not find a significant main effect of interaction format on learning performance. The absence of this main effect led us to conclude that there is no evidence in support of a near-hand effect in our Desktop-VR study with mouse interaction. In this Study 1, “hand proximity” was operationalized as holding the mouse cursor directly onto the 3D-models of the fish (object movement) rather than using the mouse on the whole virtual scene displayed on the monitor (camera movement). The results do not indicate that the mouse interaction gave the impression of directly “touching” the 3D-objects in the Desktop-VR (e.g., in terms of using the mouse as a tool to enlarge the distance that we can reach, see Brockmole et al., 2013). Furthermore, it is worth noting that participants in the camera movement condition may have used the mouse cursor next to or even on the 3D-models of the fish, potentially affecting the results.

A possible explanation for the lack of superiority of the interactive conditions compared to the no interaction condition could be that the restricted perspective used in the no interaction condition might already have been very well chosen. This perspective was selected for prior studies by a domain expert, and it might have been well suited for learning about fish movement patterns. Allowing participants to interact with the 3D-objects or the virtual scene to change the perspective might have led to masking the relevant movements. In this domain, the saliency of visual details and movements does not necessarily indicate how relevant a certain aspect is for recognizing the movement pattern relevant for propulsion (see also Lowe, 2003), so it might be disadvantageous for the learners to orient the 3D-models or the virtual scene according to more salient aspects of the 3D-models or the depicted dynamics.

In relation to learners’ visuospatial ability, our findings align with our hypothesis H2, demonstrating that learners with higher visuospatial ability outperformed those with lower visuospatial ability across all learning outcome measures. This finding is consistent with previous research on learning with dynamic visualizations (e.g., Höffler, 2010), 3D-models (e.g., Huk, 2006), and virtual environments (e.g., Lee & Wong, 2014; Sun et al., 2019), as well as studies specifically focusing on classifying fish movement patterns (e.g., Imhof et al., 2011).

For the difficult recognition task, we observed an interesting interaction between learners’ visuospatial ability and the interaction format (R1) of the Desktop-VR. Accordingly, the main effect of visuospatial ability on difficult recognition has to be interpreted in the light of this significant interaction. While the object movement condition and the no interaction condition led to similar learning outcomes regardless of learners’ visuospatial ability, the camera movement condition showed a more differentiated pattern. Learners with higher visuospatial ability achieved higher results on difficult recognition items when controlling the viewpoint of the camera in the virtual scene, whereas learners with lower visuospatial ability achieved worse results on the difficult recognition items when being allowed to control the camera viewpoint. This finding suggests that visuospatial ability plays a crucial role in learning in VR environments (e.g., Sun et al., 2019) and moderates the effectiveness of different interaction formats. The ability-as-enhancer hypothesis (e.g., Höffler, 2010) seems to align best with this pattern: Having visuospatial ability at their disposal helped learners to learn better when they could change perspective by controlling the camera viewpoint on the invisible virtual sphere on two axes (similar to interacting with for example a real

globe), but not when they could directly manipulate the orientation of the 3D-objects using complex mouse interactions. This result pattern only occurred for the most demanding difficult items and further speaks against an effect of hand proximity (for learners with higher visuospatial abilities) during mouse interaction in Desktop-VR environments. The complex mouse interactions during object movement were very complex (not to say very unnatural in terms of an embodied perspective) because the virtual objects had to be rotated on three axes, whereas the mouse as an input device operates only in a two-dimensional space (see also Fröhlich & Plate, 2000). The learners may not have extensively utilized this complex interaction pattern (object movement), which could account for its similar outcome to the no-interaction condition.

Moreover, higher visuospatial ability prevented participants from experiencing simulator motion sickness, whereas learners with lower visuospatial ability tended to experience motion sickness particularly in the no interaction condition. This indicates that visuospatial ability might play a compensating role regarding motion sickness during different interaction formats. Regarding the relationship between motion sickness and the level of interaction, often referred to as navigational control within an application, it has been shown that greater control over the environment can lead to reduced simulator sickness (e.g., Stanney & Hash, 1998). Further research has explored the occurrence of motion sickness (e.g., Boletsis & Cedergren, 2019; Coomer et al., 2018; Rebenitsch & Owen, 2016). However, a common limitation among these studies is that visuospatial ability as an individual factor is often neglected. In the present study, the higher values of motion sickness did not result in worse learning outcomes for learners with lower visuospatial ability in the no interaction condition. Nonetheless, the possibility of having control over the Desktop-VR might play a stronger role for learners who lack visuospatial abilities.

Familiarity with the domain did not foster learning but helped participants to feel more present in the Desktop-VR environment. Moreover, both interaction conditions led to higher presence than the no interaction condition: camera movement led to higher spatial presence and object movement at least tended to lead to higher ratings on the general presence item.

In conclusion, Study 1 served as a first investigation into the impact of different interaction formats (object movement, camera movement, no interaction) and learners' visuospatial ability as well as its moderating role during learning (recognition, factual knowledge) in a low immersive desktop "VR" environment. The results shed light on the complexities of interaction design and learners' characteristics in VR learning environments. Building upon the insights from Study 1, Study 2 was designed to further explore these factors in an immersive VR environment, adapting the methodological approach from Study 1 for the two interaction formats, respectively.

Study 2

In the following, we explain how Study 2 differs from Study 1. It is important to note that the essential measurements (e.g., recognition and factual knowledge, as well as visuospatial ability) in Study 2 are identical to Study 1 and are therefore not reported again.

Method

Participants and Design

In Study 2, 65 participants (19 males, 46 females, 0 diverse; 54 right-handed, 11 left-handed, 0 ambidextrous; 59 university students, 1 employed professional, 5 identified as either job seekers, unemployed, pupils, or housewives/husbands) with an average age of 25.18 years ($SD = 4.96$) were recruited via a participant recruiting system (www.iwm.sona-systems.com; selection criteria: native German speaker, age: 18–40 years). Participants were reimbursed 10 € for approximately 60 min. The study was approved by the ethics committee of the IWM (LEK 2020/043). Participants were randomly assigned to one of two conditions in a between-subject design investigating the influence of the first independent factor *interaction format* with the factor levels: first, object movement ($n = 33$) and second, viewer movement ($n = 32$; this was termed camera movement in Study 1). In contrast to Study 1, this study was not conducted in a Desktop-VR in an online setting, but in an immersive VR in a controlled laboratory environment. Thus, we operationalized the two interaction formats according to the used immersive VR: all participants conducted the study with a VR headset and controller in an immersive VR. Object movement was implemented via VR controller input and viewer movement was implemented by allowing participants to walk and move in the VR environment. Moreover, participants' *visuospatial ability* was assessed as a second continuous independent factor.

Measures

Dependent variables were the same measures as in Study 1 (easy, medium, and difficult recognition, factual knowledge, presence, and motion sickness).

Materials and Environment

Within the immersive VR learning environment, participants were exposed to the same highly realistic 3D-models of fish with the same instruction (i.e., “study the 3D-models to learn how to classify fish based on their movement patterns”) and the same additional verbal explanations used in Study 1. However, in Study 2, the virtual learning environment was presented to the participants through a VR headset (HTC Vive Pro V1), which allows a high degree of immersive virtual reality (Villena-Taranilla et al., 2022). The HTC Vive Pro allowed the participants to experience stereoscopic vision and controller interaction capabilities. Each eye had a



Fig. 7 Screenshot of the participants' view in the object movement condition

resolution of 1440×1600 pixels, and the field of view was 110° . The test room was equipped with two base stations that allowed the participants to move around with a significant degree of freedom. Participants used the associated HTC Vive Pro controllers as an input device.

Interaction Format

The two different interaction formats *object movement* and *viewer movement* were implemented as follows. In the *object movement condition*, participants were instructed to remain in a fixed position within the real room and thereby also in the immersive VR environment and not step outside. Participants were allowed to interact with the virtual 3D-models of fish to change their perspective on the fish using the controller (objects were rotated, viewers stayed at their position). Within the VR environment, a virtual hand representing the controller was displayed (see Fig. 7). In the *viewer movement condition*, participants were allowed to walk and move around (e.g., bend down to look from below at) the displayed 3D-models of the fish without any visible avatar. However, in this condition, participants were not allowed to touch or rotate the 3D-objects (objects stayed in their position, viewers moved in the virtual scene). The available space for movement within the virtual environment was approximately $3 \text{ m} \times 2.85 \text{ m}$.

Procedure

In contrast to Study 1, participants in Study 2 did not participate online in the study but attended individual appointments in our VR lab. When arriving in the lab, participants were welcomed, had to read the general information about the study, and gave their informed consent. Study 2 comprised the preliminary phase, the learning phase, the assessment of additional variables, and the testing phase and the

Table 3 Means and standard errors for easy, medium, and difficult recognition and factual knowledge (in % correct) for higher and lower visuospatial ability (VSA) in both experimental conditions of Study 2

	Object movement ($n = 33$)				Viewer movement ($n = 32$)			
	Higher VSA		Lower VSA		Higher VSA		Lower VSA	
	M	SE	M	SE	M	SE	M	SE
Easy	76.51	6.67	49.14	7.71	55.02	8.16	63.98	7.18
Medium	72.17	5.42	50.72	6.26	64.91	6.63	66.93	5.84
Difficult	59.72	3.75	48.28	4.34	62.44	4.59	55.26	4.04
Factual knowledge	74.88	2.95	60.91	3.41	68.61	3.61	65.57	3.18

procedure was as comparable as possible to the one used in Study 1. Participants answered the questionnaires from the preliminary phase (demographics, familiarity with the domain of fish movements, PFT, Ekstrom et al., 1976) on a PC. Following the preliminary phase, participants were instructed about the respective interaction format in the immersive VR depending on their experimental condition: either having the option to rotate the object physically via the controller (object movement) or having the option to walk and move around the object (viewer movement). Subsequently, participants were immersed into the VR experience by putting on the VR headset and the learning phase started with the submarine exercise, before they read the written introduction to “underwater locomotion”. Then, the four 3D-models of the to-be-learned movement patterns were presented in the same manner as in Study 1 but this time in the VR headset in the immersive VR environment. Following the learning phase, participants removed the VR headset and had to answer the VRSQ (Kim et al., 2018) and the IPQ (Schubert et al., 2001), before they answered the learning outcome measures in the testing phase on a PC. Finally, participants were informed about the study goals and compensated for their participation.

Results

Regarding the comparability between the experimental conditions, we found no significant differences (all $p_s > .323$) between the two groups (see Appendix Table A2 for means and values) in participants’ age, gender, handedness, occupation, familiarity with the domain, or visuospatial ability.

Learning Outcomes

Recognition performance in terms of easy, medium, and difficult recognition, as well as factual knowledge was analyzed with ANCOVAs with the factor interaction format (object movement, camera movement) and the continuous factor learners’ visuospatial ability (z -standardized with interaction term) as well as learners’ familiarity with the domain as a covariate (without interaction terms; see Table 3 for adjusted means and standard errors). Higher visuospatial ability was defined as

mean plus one standard deviation, whereas lower visuospatial ability was defined as mean minus one standard deviation.

There was no significant main effect of interaction format on easy recognition, medium recognition, difficult recognition, and factual knowledge (all $ps > .245$). Contrary to our hypothesis H1, we observed no overall positive effect of direct manipulation in the object movement condition compared to the viewer movement condition on learning outcomes.

Regarding learners' visuospatial ability, there was a significant main effect on difficult recognition ($F(1, 60) = 4.916$, $MSE = 267.881$, $p = .030$, $\eta^2_p = .076$) as well as on factual knowledge ($F(1, 60) = 6.631$, $MSE = 165.928$, $p = .013$, $\eta^2_p = .100$), but not for easy and medium recognition (both $ps > .113$). Participants with higher visuospatial ability achieved higher performance for difficult items and factual knowledge than those learners with lower visuospatial ability. This partially supports our hypothesis H2, that higher visuospatial ability positively influences learning performance as compared to lower visuospatial ability.

Moreover, there was a significant interaction between interaction format and learners' visuospatial ability for easy recognition performance ($F(1, 60) = 5.524$, $MSE = 846.796$, $p = .022$, $\eta^2_p = .084$; see Fig. 8). Bonferroni-adjusted pairwise comparisons revealed that participants with higher visuospatial ability profited from object movements compared to viewer movements ($p = .046$) within easy recognition tasks. In contrast, there was no difference between the two interaction formats for learners with lower visuospatial ability ($p = .172$).

Furthermore, there was a (non-significant) tendency for an interaction between interaction format and learners' visuospatial ability for medium recognition performance ($F(1, 60) = 3.481$, $MSE = 558.674$, $p = .067$, ns, $\eta^2_p = .055$; see Fig. 9). Bonferroni-adjusted pairwise comparisons revealed that participants with lower

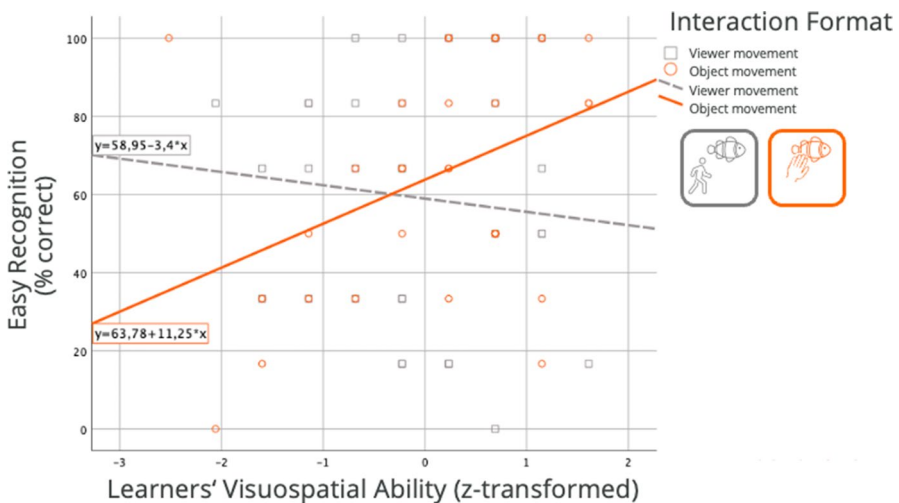


Fig. 8 Interaction between interaction format and learners' visuospatial ability on easy recognition in Study 2

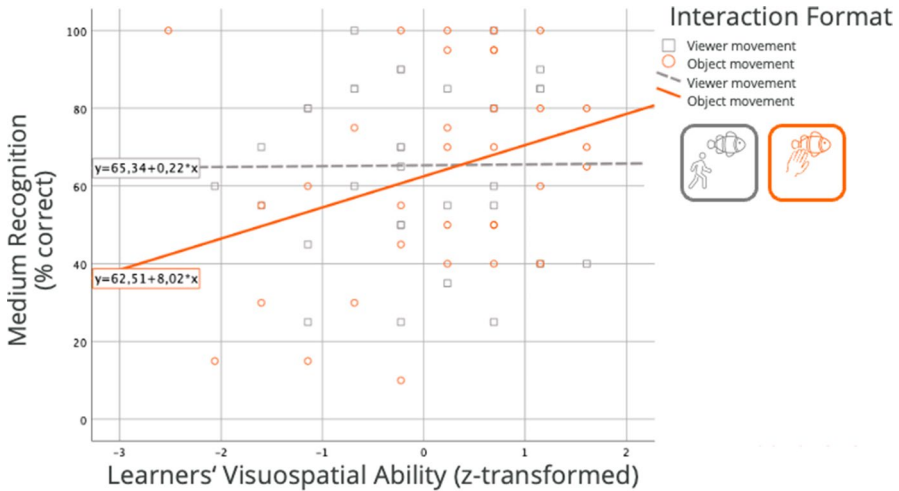


Fig. 9 (Non-significant) tendency for an interaction between interaction format and learners' visuospatial ability on medium recognition in Study 2

visuospatial ability tended to suffer from object movements compared to viewer movements ($p = .068$), within medium tasks. In contrast, there was no difference between the two interaction formats for learners with higher visuospatial ability ($p = .405$).

There was a main effect of the covariate familiarity with the domain for medium recognition ($F(1, 60) = 4.961$, $MSE = 558.674$, $p = .030$, $\eta^2_p = .076$) and factual knowledge ($F(1, 60) = 4.500$, $MSE = 165.928$, $p = .038$, $\eta^2_p = .070$). Higher familiarity with the domain led to better performance for medium recognition and factual knowledge. There were no other significant main effects or interactions concerning learning outcome measures (all $ps > .109$).

Motion Sickness and Presence

Motion sickness, as well as the four subcategories of presence (general presence item, spatial presence, involvement, and realness) were again analyzed with ANCOVAs with the factor interaction format (object movement versus camera movement) and the second continuous factor learners' visuospatial ability (z-standardized with interaction term) as well as learners' familiarity with the domain as a covariate (without interaction terms; see Table 4 for adjusted means and standard errors).

There were no main effects or interactions on motion sickness in Study 2 (all $ps > .119$). Regarding presence, there was a (non-significant) tendency for a main effect of interaction format on the general presence item ($F(1,60) = 3.952$, $MSE = 1.932$, $p = .051$, ns, $\eta^2_p = .062$). Participants in the viewer movement condition tended to rate the general presence item higher than participants in the object movement condition. In addition to observing a (non-significant) tendency towards a main effect, a significant interaction was found between interaction format and visuospatial ability regarding the general presence item ($F(1,60) = 10.297$, $MSE = 1.932$,

Table 4 Means and standard errors (in parentheses) for motion sickness, general presence item, spatial presence, involvement, and realism for higher and lower visuospatial ability (VSA) in both experimental conditions of Study 2

	Object movement (<i>n</i> = 33)				Viewer movement (<i>n</i> = 32)			
	Higher VSA		Lower VSA		Higher VSA		Lower VSA	
	M	SE	M	SE	M	SE	M	SE
Motion sickness (1–5)	1.69	0.13	1.98	0.15	1.63	0.16	1.81	0.14
General presence (1–7)	4.89	0.32	3.85	0.37	4.40	0.39	5.73	0.34
Spatial presence (1–7)	5.12	0.23	4.41	0.27	4.78	0.28	5.55	0.25
Involvement (1–7)	4.60	0.29	3.92	0.33	3.99	0.35	4.22	0.31
Realness (1–7)	3.08	0.26	3.18	0.30	3.20	0.32	2.97	0.28

$p = .002$, $\eta^2_p = .146$; see Fig. 10). Bonferroni-adjusted pairwise comparisons revealed that only participants with lower visuospatial ability rated general presence higher in the viewer movement condition than in the object movement condition ($p < .001$), whereas there was no difference between the two conditions for participants with higher visuospatial ability ($p = .339$).

Moreover, there was a significant interaction between interaction format and visuospatial ability on spatial presence ($F(1,60) = 7.868$, $MSE = .999$, $p = .007$, $\eta^2_p = .116$; see Fig. 11). Bonferroni-adjusted pairwise comparisons revealed (the almost identical pattern as for the general presence item) that only participants with lower visuospatial ability rated spatial presence higher in the viewer movement condition than in the object movement condition ($p = .003$), whereas there was again no

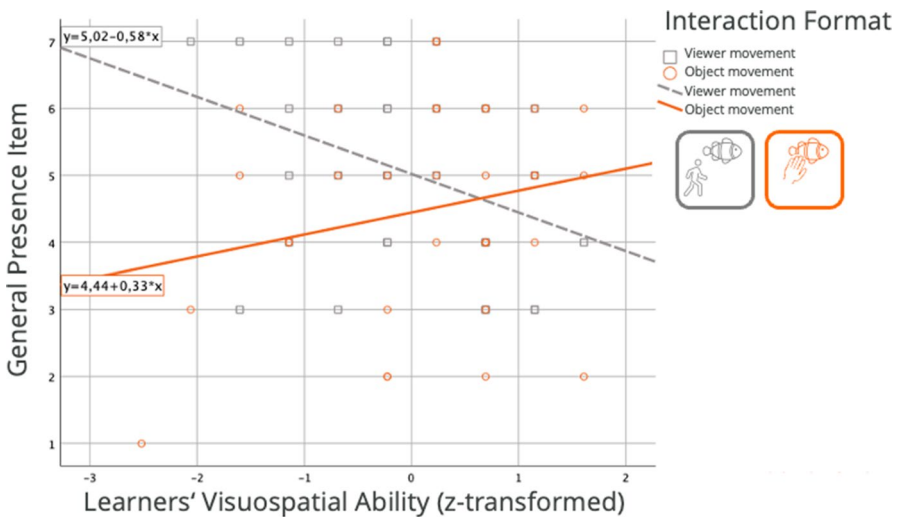


Fig. 10 Interaction between interaction format and learners' visuospatial ability on the general presence item in Study 2

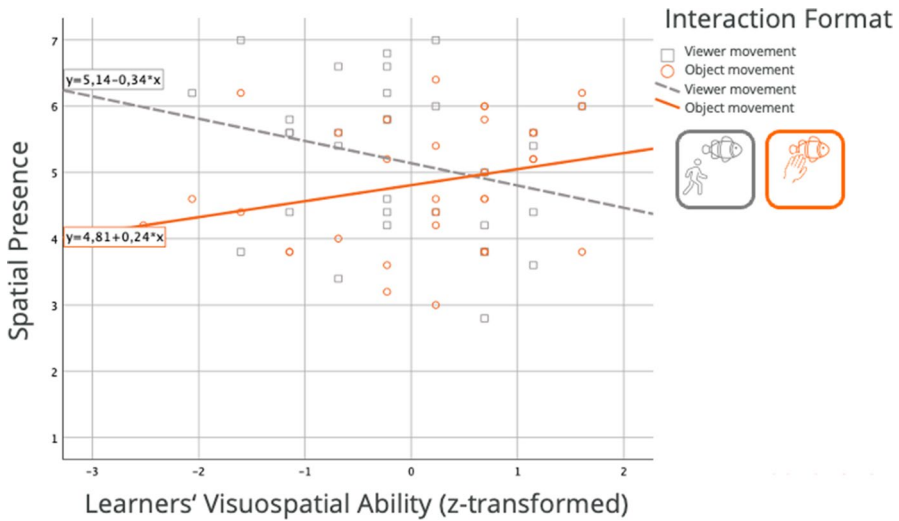


Fig. 11 Interaction between interaction format and learners' visuospatial ability on spatial presence in Study 2

difference between the two conditions for participants with higher visuospatial ability ($p = .351$).

Furthermore, there was a main effect of familiarity with the domain on the general presence item ($F(1,60) = 7.387$, $MSE = 1.932$, $p = .009$, $\eta^2_p = .110$) and spatial presence ($F(1,60) = 4.996$, $MSE = .999$, $p = .029$, $\eta^2_p = .077$) in the direction that higher familiarity with the domain led to higher ratings for the general presence item and spatial presence. There were no other main effects or interactions concerning the general presence item, spatial presence, involvement, and realism (all $ps > .115$).

Discussion of Study 2

In the second study, we investigated the influence of different interaction formats (object movement, viewer movement) on learning performance in a real immersive VR setting. Our initial hypothesis H1 suggested that object movement would lead to better learning outcomes than viewer movement because during object manipulation the virtual 3D-objects in the VR environment are presented near the (real) hands of the users and thus should be readily considered for immediate interaction and action (e.g., Gozli et al., 2012), thereby making available additional cognitive resources for learning (e.g., Agostinho et al., 2016). Moreover, we hypothesized in H2, based on prior research (e.g., Höffler, 2010), that learners with higher visuospatial ability would outperform those with lower visuospatial ability across all learning outcome measures.

Regarding hypothesis H1 on interaction format, we were not able to find general evidence in terms of a superiority of the object movement condition. Nevertheless, the significant interaction between interaction format and visuospatial ability (R1)

on easy recognition revealed that learners with higher visuospatial ability achieved better results on these easy recognition items when they were allowed to directly manipulate the virtual objects manually instead of walking around them in the virtual scene. This result pattern on easy recognition can be interpreted as an effect of hand proximity (e.g., Brockmole et al., 2013; Brucker et al., 2021; Reed et al., 2006; Tseng et al., 2012) for learners with higher visuospatial ability. The finding that this positive effect of hand proximity only occurs for the easy recognition in our study might be interpreted in terms of correspondence of these test items to the task focus and the respective task demands (see Goodhew & Clarke, 2016; Liepelt & Fischer, 2016). In prior studies on hand proximity, we could show that the influence of hand proximity effects is particularly pronounced in tasks that align closely with the instructional intent and the task requirements (e.g., Brucker et al., 2017; Weber, 2016). We found positive effects of hand proximity for a task on memorizing different dynamic sequences when learning about mitoses, whereas we found positive effects of hand proximity for a task on recognizing specific movements when learning about dance steps (it should be noted that in both studies—mitosis and dance steps—there were tasks on both memorizing different dynamic sequences as well as recognizing specific movements). Moreover, in the domain of learning about dance steps, that is visually very fast and highly complex and thereby a very similar task to the present one of learning about biological movements, we found that particularly learners with higher visuospatial ability profited from hand proximity (e.g., Brucker et al., 2017; Weber, 2016). The easy items in the present study correspond at the most with the dynamics depicted repeatedly in the 3D-models during the learning phase, because they only show relevant information. In the learning phase, there was a strong focus on visualizing the correct movements for propulsion, but not to discriminate propulsive movements from navigational ones or to identify partly invisible movement patterns. Therefore, the easy items can be seen more as pure perceptual retrieval exactly fitting the present task demands, whereas the medium and particularly the difficult items (as they show more irrelevant information or only parts of the relevant information) can be interpreted as being farther away from the present tasks demands and thereby going more in the direction of transferring knowledge based on inferences to new situations and challenges. Particularly, learners with higher visuospatial ability seem to profit from hand proximity during recognizing movement patterns. Learners with higher visuospatial ability might have more general competencies to process complex dynamic phenomena, and their advanced imagination may enable a more accurate evaluation of how a dynamic 3D-model can be rotated to achieve an optimal viewing position and thus they might be able to use the additional attentional or cognitive resources given by hand proximity more effectively (see ability-as-enhancer assumption, Höffler, 2010).

However, the descriptive result pattern and the tendency for an interaction between interaction format and visuospatial ability on medium recognition, indicated a potential inverse relationship to the hand proximity effect during direct object manipulation in immersive VR for individuals with lower visuospatial ability as they tended to profit from viewer movement instead of object movement.

Our research supports Hypothesis H2, showing that participants with higher visuospatial ability perform better than those with lower visuospatial ability in several

Table 5 Means and standard errors for easy recognition, medium recognition, difficult recognition, and factual knowledge (in % correct). Absolute values for motion sickness, general presence item, spatial presence, involvement, and realness for low and high immersive learning environments

	Low immersive Desktop-VR; N=160		High immersive immersive VR; N=65	
	M	SE	M	SE
Easy recognition	68.12	27.11	62.31	30.23
Medium recognition	69.91	19.31	64.39	24.58
Difficult recognition	56.84	16.24	56.52	16.75
Factual knowledge	69.49	12.01	67.84	13.73
Motion sickness (1–5)	1.59	0.55	1.78	0.58
General presence item (1–7)	2.84	1.46	4.79	1.53
Spatial presence (1–7)	3.29	1.10	5.01	1.06
Involvement (1–7)	3.15	1.23	4.21	1.24
Realness (1–7)	2.45	0.95	3.10	1.12

areas. These areas include difficult recognition tasks (in general across both interaction formats) and factual knowledge tasks as well as medium and easy recognition tasks in the object movement condition (see the respective interaction patterns). With these observed interaction patterns, the data also gives evidence for research question R1 regarding the possible interaction between visuospatial ability and different interaction formats. This clearly demonstrates the influence of participants' characteristics in terms of their visuospatial ability during learning about movements in VR environments. In both Study 1 and Study 2, greater familiarity with the domain resulted in higher presence ratings. Moreover, in Study 2, greater familiarity with the domain also contributed to improved learning outcomes (for medium recognition and factual knowledge).

However, there were no differences for motion sickness observable in the immersive VR environment. This is still in line with the results of Study 1, as we compared in Study 2 the two interactive conditions, but not the no interaction condition, that was accountable for this effect for learners with lower visuospatial ability in Study 1. Furthermore, participants with lower visuospatial ability experienced higher presence in the viewer movement condition than in the object movement condition, whereas there was no such difference between the two conditions for participants with higher visuospatial ability. This was in line with the result pattern on learning outcomes, as learners with lower visuospatial ability tended to learn better in the viewer movement condition.

Comparison Between Study 1 and 2

In an additional exploratory step, we directly compared learning outcomes (in terms of easy, medium, and difficult recognition, as well as factual knowledge), motion sickness, and presence for the two studies with independent samples t-tests for the factor immersion (low = Desktop-VR versus high = immersive VR; see Table 5 for

means and standard deviations as well as Tables 6 and 7 for correlations between the dependent variables in Study 1 and Study 2, respectively).

Learning Outcomes

There were no significant differences between the two VR environments for all learning outcome variables: easy recognition (95%-CI[-2.311, 13.946], $t(223)=1.410$, $p=.160$, ns, $d=0.207$); a (non-significant) tendency for medium recognition (95%-CI[-.553, 11.596], $t(223)=1.791$, $p=.075$, ns, $d=0.263$); and difficult recognition (95%-CI[-4.426, 5.074], $t(223)=-.134$, $p=.447$, ns, $d=0.020$), as well as factual knowledge (95%-CI[-1.976, 5.287], $t(223)=.898$, $p=.185$, ns, $d=0.132$).

Motion Sickness and Presence

There was a significant difference for all variables in the direction that the higher immersive VR resulted in higher ratings on motion sickness (95%-CI[-.354, -.021], $t(223)=-2.276$, $p=.012$, $d=-0.335$), as well as on the general presence item (95%-CI[-2.376, -1.518], $t(223)=-8.942$, $p<.001$, $d=-1.315$); on spatial presence (95%-CI[-2.036, -1.404], $t(223)=-10.733$, $p<.001$, $d=-1.578$); on involvement (95%-CI[-1.421, -.708], $t(223)=-5.875$, $p<.001$, $d=-0.865$); and on realism (95%-CI[-.936, -.356], $t(223)=-4.394$, $p<.001$, $d=-0.646$).

The direct comparison between the two studies indicated that the immersive VR in Study 2 yielded more presence and higher motion sickness but did not lead to better learning outcomes than the Desktop-VR in Study 1 (for medium recognition the values tended to be even higher in the Desktop-VR than in the immersive VR environment). This must be considered when deciding for a specific type of VR environment during designing learning environments.

General Discussion

In two studies, we investigated how the possibility of interacting with the perspective (object movement versus camera/viewer movement) of virtual 3D-models in a low (Study1: desktop) and high (Study 2: headset) immersive VR environment influences learning to classify different movement patterns. Beyond the possibility of interacting and the degree of immersion in the VR, we addressed participants' visuospatial ability because it plays a pivotal role in learning about dynamic movements in VR environments (e.g., Höffler, 2010; Skulmowski, 2023b).

It is crucial to highlight that our research concentrates on exploiting materials that demonstrate complex dynamic movements as the focal point of investigation. This aspect represents a significant strength of our research, particularly when contrasted with other studies in this domain, which often concentrate on knowledge acquisition derived from static (e.g., Jang et al., 2017) or less complex subjects.

Table 6 Correlation matrix for dependent and additional variables in Study 1 (low immersive learning environment = Desktop-VR; $N = 160$)

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Easy recognition	—										
2. Medium recognition	.63**	—									
3. Difficult recognition	.44**	.71**	—								
4. Factual knowledge	.43**	.52**	.42**	—							
5. General presence	-.01	-.04	-.07	.03	—						
6. Spatial presence	-.11	-.06	-.07	-.07	.72**	—					
7. Involvement	.00	.07	.05	.04	.48**	.44**	—				
8. Realness	-.09	-.02	-.14	-.06	.62**	.53**	.46**	—			
9. Motion sickness	-.12	-.05	-.03	-.13	-.24**	-.21**	-.04	-.07	—		
10. Domain familiarity	.03	-.08	-.08	-.03	.23**	.15	-.07	.07	-.04	—	
11. Visuospatial ability (PFT)	.16*	.17*	.25**	.24**	.09	.07	.04	.03	-.19*	-.05	—

* $p < .05$; ** $p < .01$

Table 7 Correlation matrix for dependent and additional variables in Study 2 (high immersive learning environment = immersive VR; $N = 65$)

Variable	1	2	3	4	5	6	7	8	9	10	11
1. Easy recognition	—										
2. Medium recognition	.72**	—									
3. Difficult recognition	.49**	.69**	—								
4. Factual knowledge	.54**	.66**	.56**	—							
5. General presence	-.05	.01	.00	-.07	—						
6. Spatial presence	-.13	.08	.08	-.00	.68**	—					
7. Involvement	.13	.15	.11	.25*	.30*	.45**	—				
8. Realness	-.34**	-.31*	-.27*	-.11	.39**	.41**	.28*	—			
9. Motion sickness	-.06	-.28*	-.29*	-.17	-.07	-.16	-.00	-.07	—		
10. Domain familiarity	.11	.18	.11	.17	.20	.16	.09	.17	-.10	—	
11. Visuospatial ability (PFT)	.18	.18	.24	.31*	-.06	-.02	.10	.00	-.18	-.13	—

* $p < .05$; ** $p < .01$

We compared in both studies two interactive conditions: The *object movement condition* allowed participants to rotate the dynamic virtual 3D-objects (while the camera or the viewer respectively was not moved), whereas the *camera/viewer movement condition* allowed changing the viewpoint of the camera (or the viewer respectively) by moving around the virtual 3D-objects (while the virtual 3D-objects itself was not moved).

Results revealed a generally positive effect of learners' visuospatial ability (in line with hypothesis H2) on learning. Future studies should explore how visuospatial skills and familiarity with the domain (as a possible indicator of prior knowledge) influence learning efficiency because they are connected to general intelligence (e.g., Buckley et al., 2018; Carroll, 1993). This research could further examine the specific contributions of each factor to general intelligence, and to which extend participants who have these variables at their disposal, are better learners overall, as they show enhanced individual learning efficiency and capacity to manage cognitive load (e.g., Sweller, 2020). Insight into these relationships could enhance educational methods and our understanding of intelligence and learning.

Besides, the general positive effects of learners' visuospatial ability, our results indicate that it also moderates the effectiveness of different interaction formats in VR environments (in line with research question R1). Interestingly, not only the characteristics of the learners (between-person differences, Lachmair et al., 2022) but also the type of used VR environment in terms of how immersive it was (situatedness, Lachmair et al., 2022; see also Fischer and Brugger 2011) played a role for the effectiveness of more or less embodied interaction formats in this domain. In the low immersive Desktop-VR (Study 1), learners with higher visuospatial ability profited from camera movement. In contrast, in high immersive VR (Study 2), learners with higher visuospatial ability profited from object movement, whereas learners with lower visuospatial ability tended to suffer from object movement. This interaction pattern and its possible explanations will be discussed in more detail in the following section. Although we did not find a dominant main effect for interaction format (as initially hypothesized in H1) in either study, we did find in line with research question R1 that learners' visuospatial ability moderated the effectiveness of the different interaction formats.

Practical and Theoretical Implications

In comparison, the interaction possibility in the object movement condition in the immersive VR environment of Study 2 is much more an actual "near hand" experience than the mouse cursor interaction in the Desktop-VR environment (of Study 1). This is explained by the fact that the manipulation of the virtual 3D-objects via the controllers (Study 2) is done by the real hands of the participants (with the controllers in it). The hands with the controller must be held near the virtual 3D-objects because the object cannot be manipulated if the hands are too far away. Thus, the proprioceptive information of the hands near the stimuli is given (Reed et al., 2006). Moreover, to ensure some kind of visual information, we implemented a virtual hand representing the real hand with the controller in the immersive VR environment

(effects of rubber hands, see Reed et al., 2006). However, this object manipulation was only beneficial for easy recognition items and learners with higher visuospatial ability in the immersive VR. This seems to be the combination under which potential beneficial effects of hand proximity can unfold. To the best of our knowledge, no studies directly investigate hand proximity effects on learning in VR environments (see Peck & Tutar, 2020, for a study on avatar hand proximity on working memory in VR). Thus, our study is a first promising step, and future research is needed to further disentangle hand proximity effects in different virtual learning environments in combination with the moderating role of participants' visuospatial ability. We are convinced that with the increasing importance of VR environments in the future, this effect has great potential for further research.

In addition to the rationale that object manipulation in immersive VR is advantageous due to the close proximity between the learners' hands and the virtual (dynamic) 3D objects, other complementary explanations could be the level of embodiment and agency (Johnson-Glenberg, 2018; Makransky & Petersen, 2021; Skulmowski & Rey, 2018), in terms of the degree of embodied interaction, experienced in the different interaction formats as well as the cognitive resources involved with the different experimental conditions (e.g., Sweller, 2020). Even though in both studies, the two ideas of either manipulating the virtual object or manipulating the camera's viewpoint (respectively, the viewer's viewpoint) were implemented, the resulting four variants of the expression of the implemented interaction also varied due to the different VR environments. The object movement via mouse control in the Desktop-VR was very complex and unnatural due to the rotation of the 3D-models on three axes with the mouse as an input device designed to operate only in a two-dimensional space. Thus, this interaction format is the least embodied, and participants could not rely on existing experiences or even evolved abilities and thus this format must be considered secondary knowledge (instead of primary knowledge, see Van Gog et al., 2009). At the same time, this format of object movement in the Desktop-VR might be the one exposing the most cognitive resources involved with interleaving the new information given by changing the fish angle and the new manipulation technique of rotating the virtual object on three axes with the mouse input device. In contrast, the viewer movement condition in the true immersive VR was very embodied and natural, giving haptic feedback by the feet touching the ground and using the whole body in order to make viewpoint changes. Walking around a virtual object in an immersive VR exactly works like walking around a real object (e.g., a sculpture in a museum), and the motoric behavior is the same and has not been translated via any input device, even though the visual input differs. Thus, this viewer movement is entrenched, the most grounded, embodied, and maybe even well-situated (see Fischer & Brugger, 2011), and can be considered primary knowledge (Van Gog et al., 2009), thereby building the other pole on a continuum of the level of embodiment of our four different variants of interaction formats.

The two remaining conditions can be considered being positioned between these two extremes (object movement in Desktop-VR = highly unnatural secondary knowledge versus viewer movement in immersive VR = highly natural primary knowledge). The object movement via controller in immersive VR and the camera movement via mouse interaction in the Desktop-VR can be considered more natural

than the object movement in Desktop-VR. However, at the same time, they are still less natural than the viewer movement in immersive VR as they both make use of input devices to interact with the virtual elements: the camera movement interaction in the Desktop-VR via mouse input on two axes on the invisible sphere and the object movement in the immersive VR via a controller on three axes (only using their own body parts to hold and move the controller).

In both these conditions, participants had to interact with an entity before them via an input device. In the object movement via the controller in the immersive VR, this entity is the virtual 3D-object, whereas in the camera movement via mouse interaction in the Desktop-VR condition, this entity is the invisible sphere on the computer screen that surrounds the 3D-models. To move and rotate these entities into a position that allows observing a particular perspective on the depicted dynamics of the fish model, it is necessary to build a dynamic mental model (e.g., Lowe & Boucheix, 2017), based on which the final perspective that the learner wants to see can be imagined. The change from the respective starting position of the perspective (either the fish model itself or its surrounding virtual sphere) to the desired output situation invokes the ability of mental rotation and, therefore, the visuospatial ability of the learners. This might be why participants with higher visuospatial ability could improve their learning in these two conditions. Moreover, it might be the case, that incongruences between the changes needed to view a new angle of the fish and the needed manipulation technique for the input device (either the invisible sphere or the VR controller) may have been more difficult for learners with lower visuospatial ability to integrate, whereas learners with higher visuospatial ability managed to do so.

It is essential to acknowledge that the same assumption applies to the object movement via mouse in the Desktop-VR condition. However, our experience from Study 1 indicates that rotating a virtual object with the mouse on three axes is considerably more challenging than rotating the same object with the controller held in hand on three axes. Additionally, it is more complex than rotating a similar virtual object with the mouse on two axes (familiar mouse control, see Smith, 2019). This is because there is no intuitive mapping between the to-be-executed movements in the two-dimensional mouse operation space and the imagined dynamic mental model transformations on the three rotation axes.

In sum, our two studies highlighted the importance of taking learners' visuospatial ability into account (e.g., Skulmowski, 2023b) and that results on interaction formats in Desktop-VR environments are not transferable on a one-to-one basis to those in true immersive VR environments. Thus, during designing virtual learning environments, not only implementing bodily engagement in terms of interaction possibilities (cf. Johnson-Glenberg, 2018; Skulmowski & Rey, 2018), but also considering participants' learner characteristics (e.g., Höffler, 2010) as well as the properties of the different environmental situations are essential and must be considered (e.g., Lachmair et al., 2022).

Limitations and Future Research

One area for potential improvement in future studies is to enhance the level of agency (Johnson-Glenberg, 2018) or bodily engagement and task integration, particularly concerning the taxonomy of embodiment in education, as Skulmowski and Rey (2018) proposed. However, when considering agency, it is important to note that it may not always be optimal to allow participants complete freedom of choice; people need to understand what they are doing and why. The perspectives we offered were selected based on prior studies by a domain expert, and thus they have with a high likelihood been the best possible perspective for learning about fish movement patterns. Therefore, allowing participants to interact with the 3D models or the virtual scene to change perspectives might have obscured relevant movements. Another aspect that could be further explored is whether full agency or educated scaffolding (perhaps with ways of interaction) is more effective for learning in VR (see Doo et al., 2020). This should be investigated further in future VR research, as we did not examine which positions or viewpoints were chosen and viewed for how long by the participants (both participants performing rather low as well as participants performing rather high on learning outcomes).

In Study 1, bodily engagement was limited as participants conducted the study while seated. In contrast, Study 2 involved higher bodily engagement, as participants were encouraged to stand and move around during the learning task. However, both studies still employed an incidental approach, with movement not directly integrated into the learning process. Moreover, the movements were not only not aligned with the depicted content (task integration, Skulmowski & Rey, 2018), but they were also not aligned with the underlying processes needed for comprehension (e.g., predicting or imagining the dynamics, see De Koning & Tabbers, 2011; Lowe & Boucheix, 2017). To achieve a higher level of bodily engagement and task integration, future research could, for instance, require participants to perform specific bodily movements or gestures to imitate locomotion patterns (e.g., De Koning & Tabbers, 2013; Scheiter et al. 2020). The maximum potential for investigating the relationship between bodily engagement, task integration, and learning outcomes could be achieved by doing so. Such an approach would provide valuable new insights into the benefits of incorporating physical movements directly into the learning process and its impact on learning.

Moreover, we acknowledge that the findings from the immersive VR study should be replicated with a larger participant sample and include a non-interaction condition as a control condition to strengthen the study's validity. Particularly in the light of the (non-significant) tendencies observed across both studies, a replication effort that aims for higher statistical power, possibly through a within-subject study design, should be considered to clarify and solidify these findings.

Furthermore, on the one hand, we did not directly investigate which perspectives are the best ones to learn about the depicted dynamics (see Keehner et al. 2008), and on the other hand, it might also be of interest to address how much participants interacted in the different interaction formats as well as which perspectives they observed (e.g., using eye-tracking, cf. Drai-Zerbib et al., 2022), how long they observed the different perspectives, and how often they changed between the perspectives (see

Boucheix et al., 2018) as well as how these variables connect to learning outcomes. Therefore, future research, which for example records and evaluates interactions in the VR environments, is needed as a feasible step to disentangle these effects. Furthermore, a next interesting step would be to investigate which brain areas are increasingly activated during the different interaction formats for learners with different amounts of visuospatial ability by means of electroencephalography (EEG, e.g., Scharinger, 2018) or by means of functional near-infrared spectroscopy (fNIRS, non-intrusive method of optical imaging, e.g., Brucker et al., 2015) to shed light onto the moderating role of learners' visuospatial ability during learning in VR environments.

Conclusion

In conclusion, the two studies highlighted the importance of considering participants' visuospatial ability and the influence of different low and high immersive VR environments on the effectiveness of different interaction formats. It provided valuable insights into designing virtual learning environments that maximize learning outcomes based on learners' characteristics and environmental settings. Future research in this field should explore the integration of bodily movements in learning tasks and continue to investigate the effects of learners' characteristics and hand proximity effects in different VR environments on learning processes.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-024-09895-w>.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The data that support the findings of this study are available from the corresponding author, [B. B.], upon reasonable request.

Declarations

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agostinho, S., Ginns, P., Tindall-Ford, S., Mavilidi, M., & Paas, F. (2016). 'Touch the screen': Linking touch-based educational technology with learning – a synthesis of current research. In L. Lin & R. Atkinson (Eds.), *Educational technologies: Challenges, applications and learning outcomes* (pp. 33–57). Nova Publishing.
- Al Amri, A. Y., Osman, M. E., & Al Musawi, A. S. (2020). The effectiveness of a 3D-virtual reality learning environment (3D-VRLE) on the Omani eighth grade students' achievement and motivation towards physics learning. *International Journal of Emerging Technologies in Learning (online)*, 15(5), 4.
- Ayres, P., Marcus, N., Chan, C., & Qian, N. (2009). Learning hand manipulative tasks: When instructional animations are superior to equivalent static representations. *Computers in Human Behavior*, 25, 348–353. <https://doi.org/10.1016/j.chb.2008.12.013>
- Boletsis, C., & Cedergren, J. E. (2019). VR Locomotion in the New Era of Virtual Reality: An Empirical Comparison of Prevalent Techniques. *Advances in Human-Computer Interaction*, 7420781. <https://doi.org/10.1155/2019/7420781>
- Brown, L. E., Marlin, M. C., & Morrow, S. (2015). On the contributions of vision and proprioception to the representation of hand-near targets. *Journal of Neurophysiology*, 113, 409–419. <https://doi.org/10.1152/jn.00005.2014>
- Brucker, B., Ehllis, A.-C., Häußinger, F. B., Fallgatter, A. J., & Gerjets, P. (2015). Watching corresponding gestures facilitates learning with animations by activating human mirror-neurons: An fNIRS study. *Learning and Instruction*, 36, 27–37. <https://doi.org/10.1016/j.learninstruc.2014.11.003>
- Brucker, B., Brömme, R., Weber, S.-C., & Gerjets, P. (2017). Learning on multi-touch devices: Is directly touching dynamic visualizations helpful? 17th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI). *Special Interest Group (SIG) 2*. Tampere, Finland
- Brucker, B., Brömme, R., Ehrmann, A., Edelmann, J., & Gerjets, P. (2021). Touching digital objects directly on multi-touch devices fosters learning about visual contents. *Computers in Human Behavior*, 119, 106708. <https://doi.org/10.1016/j.chb.2021.106708>
- Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, 101, 150–167. <https://doi.org/10.1016/j.compedu.2016.06.005>
- Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, 23, 638–663. <https://doi.org/10.1002/acp.1473>
- Boucheix, J.-M., Gauthier, P., Fontaine, J. B., & Jaffeux, S. (2018). Mixed camera viewpoints improve learning medical hand procedure from video in nurse training? *Computers in Human Behavior*, 89, 418–429. <https://doi.org/10.1016/j.chb.2018.01.017>
- Bricken, M. (1991). Virtual reality learning environments: Potentials and challenges. *ACM SIGGRAPH Computer Graphics*, 25, 178–184. <https://doi.org/10.1145/126640.126657>
- Brockmole, J. R., Davoli, C. C., Abrams, R. A., & Witt, J. K. (2013). The world within reach: Effects of hand posture and tool use on visual cognition. *Current Directions in Psychological Science*, 22, 38–44. <https://doi.org/10.1177/0963721412465065>
- Buckley, J., Seery, N., Cauty, D., & Gumaelius, L. (2018). Visualization, inductive reasoning, and memory span as components of fluid intelligence: Implications for technology education. *International Journal of Educational Research*, 90, 64–77. <https://doi.org/10.1016/j.ijer.2018.05.007>
- Buehler, K., & Kohne, A. (2019). Lernen mit Virtual Reality: Chancen und Möglichkeiten der digitalen Aus- und Fortbildung. In M. Groß, M. Müller-Wiegand, & D. F. Pinnow (Eds.), *Zukunftsfähige Unternehmensführung* (pp. 209–224). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-662-59527-5_11
- Burkhardt, D., Stab, C., Nazemi, K., Breyer, M., & Fellner, D. W. (2010). Approaches for 3d-visualizations and knowledge worlds for exploratory learning. In L. G. Chova, D. M. Belenguer, & I. C. Torres (Eds.), *Edulearn10: International conference on education and new learning technologies* (pp. 6427–6437). Valenica: Iated-Int Assoc Technology Education & Development. <https://doi.org/10.5281/zenodo.3455879>
- Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. *Cambridge University Press*. <https://doi.org/10.1017/CBO9780511571312>

- Christopoulos, D., Mavridis, P., Andreadis, A., & Karigiannis, J. N. (2011). Using virtual environments to tell the story: "The Battle of Thermopylae". In *2011 Third International Conference on Games and Virtual Worlds for Serious Applications* (pp. 84–91). IEEE. <https://doi.org/10.1109/vs-games.2011.18>
- Coomer, N., Bullard, S., Clinton, W., & Williams-Sanders, B. (2018). Evaluating the effects of four VR locomotion methods: joystick, arm-cycling, point-tugging, and teleporting. In *Proceedings of the 15th ACM symposium on applied perception* (pp. 1–8). <https://doi.org/10.1145/3225153.3225175>
- Cromley, J. G., Chen, R., & Lawrence, L. (2023). Meta-analysis of STEM learning using virtual reality: Benefits across the board. *Journal of Science Education and Technology*, *32*, 355–364. <https://doi.org/10.1007/s10956-023-10032-5>
- 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*, 272–309. <https://doi.org/10.1080/15213269.2015.1015740>
- Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, *41*, 10–32. <https://doi.org/10.1111/j.1467-8535.2009.01038.x>
- Davoli, C. C., Brockmole, J. R., & Goujon, A. (2012). A bias to detail: How hand position modulates visual learning and visual memory. *Memory & Cognition*, *40*, 352–359. <https://doi.org/10.3758/s13421-011-0147-3>
- De Koning, B. B., Marcus, N., Brucker, B., & Ayres, P. (2019). Does observing hand actions in animations and static graphics differentially affect learning of hand-manipulative tasks? *Computers & Education*, *141*, 103636. <https://doi.org/10.1016/j.compedu.2019.103636>
- De Koning, B. B., & Tabbers, H. K. (2011). Facilitating understanding of movements in dynamic visualisations: An embodied perspective. *Educational Psychology Review*, *23*, 501–521. <https://doi.org/10.1007/s10648-011-9173-8>
- De Koning, B. B., & Tabbers, H. K. (2013). Gestures in instructional animations: A helping hand to understanding non-human movements? *Applied Cognitive Psychology*, *27*, 683–689. <https://doi.org/10.1002/acp.2937>
- Dengel, A., & Mägdefrau, J. (2021). Immersive learning predicted: presence, prior knowledge, and school performance influence learning outcomes in immersive educational virtual environments. In *Proceedings of the 6th International Conference of the Immersive Learning Research Network (iLRN)*, San Luis Obispo, CA, USA, pp. 163–170. <https://doi.org/10.23919/iLRN47897.2020.9155084>
- Doo, M. Y., Bonk, C., & Heo, H. (2020). A meta-analysis of scaffolding effects in online learning in higher education. *The International Review of Research in Open and Distributed Learning*, *21*, 60–80. <https://doi.org/10.19173/irrodl.v21i13.4638>
- Drai-Zerbib, V., Bernigaud, L., Gaston-Bellegarde, A., Boucheix, J.-M., & Baccino, T. (2022). Eye movements during comprehension in virtual reality: The influence of a change in point of view between auditory and visual information in the activation of a mental model. *Frontiers in Virtual Reality*, *3*, 874054. <https://doi.org/10.3389/frvir.2022.874054>
- Egeberg, M. C. S., Stine, L. R., Lind, S. L. R., Serubugo, S., Skantarova, D., & Kraus, M. (2016). Extending the human body in virtual reality: Effect of sensory feedback on agency and ownership of virtual wings. In *Proceedings of the 2016 Virtual Reality International Conference (VRIC '16)*. Association for Computing Machinery, New York, NY, USA, Article 30, 1–4. <https://doi.org/10.1145/2927929.2927940>
- Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Educational Testing Service.
- Fischer, M. H., & Brugger, P. (2011). When digits help digits: Spatial–numerical associations point to finger counting as prime example of embodied cognition. *Frontiers in Psychology*, *2*, 260.
- Fröhlich, B., & Plate J. (2000). The cubic mouse: A new device for three-dimensional input. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '00)*. Association for Computing Machinery, New York, NY, USA, 526–531. <https://doi.org/10.1145/332040.332491>
- Fuchs, P., Moreau, G., & Guittou, P. (2011). *Virtual reality: Concepts and technologies*. CRC Press. <https://doi.org/10.1201/b11612>
- Geary, D. (2007). Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology. In J. S. Carlson & J. R. Levin (Eds.), *Psychological perspectives on contemporary educational issues* (pp. 1–99). Greenwich, CT: Information Age.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20–25. [https://doi.org/10.1016/0166-2236\(92\)90344-8](https://doi.org/10.1016/0166-2236(92)90344-8)

- Goodhew, S. C., & Clarke, R. (2016). Contributions of parvocellular and magnocellular pathways to visual perception near the hands are not fixed, but can be dynamically altered. *Psychonomic Bulletin and Review*, 23, 156–162. <https://doi.org/10.3758/s13423-015-0844-1>
- Goodhew, S. C., Fogel, N., & Pratt, J. (2014). The nature of altered vision near the hands: Evidence for the magnocellular enhancement account from object correspondence through occlusion. *Psychonomic Bulletin & Review*, 21, 1452–1458. <https://doi.org/10.3758/s13423-014-0622-5>
- Gozli, D. G., West, G. L., & Pratt, J. (2012). Hand position alters vision by biasing processing through different visual pathways. *Cognition*, 124, 244–250. <https://doi.org/10.1016/j.cognition.2012.04.008>
- Grassini, S., & Laumann, K. (2020). Questionnaire measures and physiological correlates of presence: A systematic review. *Frontiers in Psychology*, 11, 349. <https://doi.org/10.3389/fpsyg.2020.00349>
- Graziano, M. S. A., & Gross, C. G. (1993). A bimodal map of space: Somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Experimental Brain Research*, 97, 96–109. <https://doi.org/10.1007/BF00228820>
- Graziano, M. S. A., Yap, G. S., & Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science*, 266, 1054–1057. <https://doi.org/10.1126/science.7973661>
- Graziano, M. S. A., Hu, X. T., & Gross, C. G. (1997). Visuospatial properties of ventral premotor cortex. *Journal of Neurophysiology*, 77, 2268–2292. <https://doi.org/10.1152/jn.1997.77.5.2268>
- Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education*, 8, 1–32. <https://doi.org/10.1007/s40692-020-00169-2>
- Hartmann, R. (2013). The Anne Frank House in Amsterdam: A museum and literary landscape goes virtual reality. *Journalism and Mass Communication*, 3(10), 625–644.
- Hartmann, T., Wirth, W., Vorderer, P., Klimmt, C., Schramm, H., & Böcking, S. (2015). Spatial presence theory: State of the art and challenges ahead. *Immersed in media: Telepresence theory, measurement & technology*, 115–135. https://doi.org/10.1007/978-3-319-10190-3_7
- Hegarty, M., & Waller, D. (2005). *Individual differences in spatial abilities. The Cambridge handbook of visuospatial thinking* (pp. 121–169)
- Hilfert, T., & König, M. (2016). Low-cost virtual reality environment for engineering and construction. *Visualisation in Engineering*, 4, 2. <https://doi.org/10.1186/s40327-015-0031-5>
- Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualisations—A meta-analytic review. *Educational Psychological Review*, 22, 245–269. <https://doi.org/10.1007/s10648-010-9126-7>
- Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17, 722–738. <https://doi.org/10.1016/j.learninstruc.2007.09.013>
- Huang, W., Roscoe, R. D., Craig, S. D., & Johnson-Glenberg, M. C. (2022). Extending the cognitive-affective theory of learning with media in virtual reality learning: A structural equation modeling approach. *Journal of Educational Computing Research*, 60, 807–842. <https://doi.org/10.1177/07356331211053630>
- Huk, T. (2006). Who benefits from learning with 3d models? the case of spatial ability. *Journal of Computer Assisted Learning*, 22, 392–404. <https://doi.org/10.1111/j.1365-2729.2006.00180.x>
- Imhof, B., Scheiter, K., & Gerjets, P. (2011). Learning about locomotion patterns from visualizations: Effects of presentation format and realism. *Computers & Education*, 57, 1961–1970. <https://doi.org/10.1016/j.compedu.2013.01.017>
- Imhof, B., Scheiter, K., Edelman, J., & Gerjets, P. (2012). How temporal and spatial aspects of presenting visualizations affect learning about locomotion patterns. *Learning and Instruction*, 22, 193–205. <https://doi.org/10.1016/j.learninstruc.2011.10.006>
- Jain, D., Sra, M., Guo, J., Marques, R., Wu, R., Chiu, J., & Schmandt, C. (2016). Immersive terrestrial scuba diving using virtual reality. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16). Association for Computing Machinery, New York, NY, USA, 1563–1569. <https://doi.org/10.1145/2851581.2892503>
- James, K. H., Humphrey, G. K., Vilis, T., Corrie, B., Baddour, R., & Goodale, M. A. (2002). “Active” and “passive” learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society Inc*, 34, 383–390. <https://doi.org/10.3758/BF03195466>

- Jang, S., Vitale, J. M., Jyung, R. W., & Black, J. B. (2017). Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education*, 106, 150–165. <https://doi.org/10.1016/j.compedu.2016.12.009>
- Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI*, 5, 81. <https://doi.org/10.3389/frobt.2018.00081>
- 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, 1263–1284. <https://doi.org/10.1111/jcal.12567>
- Johnson-Glenberg, M. C., Kosa, M., & O'Rourke, H. P. (2023). STEM learning, science identity and immersivity: Giant screen films comparing 2D, 3D, and dome formats including a videogame assessment. *Frontiers in Education*, 7, 1096889. <https://doi.org/10.3389/educ.2022.1096889>
- Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualisations: What matters is what you see, not whether you interact. *Cognitive Science*, 32, 1099–1132. <https://doi.org/10.1080/03640210801898177>
- Klemke, A., & Halfmann, M. (2020). *Versuchsumgebung: IWM-Study VRfish01 - 02*. Tübingen: Leibniz-Institut für Wissensmedien.
- Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, 69, 66–73. <https://doi.org/10.1016/j.apergo.2017.12.016>
- Lachmair, M., Fischer, M. H., & Gerjets, P. (2022). Action-control mappings of interfaces in virtual reality: A study of embodied interaction. *Frontiers in Virtual Reality*, 3, 976849. <https://doi.org/10.3389/frvir.2022.976849>
- Lee, E.A.-L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education*, 79, 49–58. <https://doi.org/10.1016/j.compedu.2014.07.010>
- Lee, H. (2007). Instructional design of web-based simulations for learners with different levels of spatial ability. *Instructional Science*, 35, 467–479. <https://doi.org/10.1007/s11251-006-9010-5>
- Li, Y., Huang, J., Tian, F., Wang, H., & Dai, G. (2019). Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware*, 1, 84–112. [10.3724/SP.J.2096-5796.2018.0006](https://doi.org/10.3724/SP.J.2096-5796.2018.0006)
- Liepelt, R., & Fischer, R. (2016). Task demands determine hand posture bias on conflict processing in a Simon task. *Psychonomic Bulletin & Review*, 23, 579–586. <https://doi.org/10.3758/s13423-015-0901-9>
- Lowe, R., & Boucheix, J.-M. (2017). A composition approach to design of educational animations. In: Lowe, R., Ploetzner, R. (eds) *Learning from Dynamic Visualization*. Springer, Cham. https://doi.org/10.1007/978-3-319-56204-9_1
- Lowe, R. K. (2003). Animation and learning: Selective processing of information in dynamic graphics. *Learning and Instruction*, 13, 157–176. [https://doi.org/10.1016/S0959-4752\(02\)00018-X](https://doi.org/10.1016/S0959-4752(02)00018-X)
- 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>
- Marcus, N., Cleary, B., Wong, A., & Ayres, P. (2013). Should hand actions be observed when learning hand motor skills from instructional animations? *Computers in Human Behavior*, 29, 2172–2178. <https://doi.org/10.1016/j.chb.2013.04.035>
- Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86, 389–401. <https://doi.org/10.1037/0022-0663.86.3.389>
- Mierowsky, R., Marcus, N., & Ayres, P. (2020). Using mimicking gestures to improve observational learning from instructional videos. *Educational Psychology*, 40, 550–569. <https://doi.org/10.1080/01443410.2019.1650896>
- Peck, T. C., & Tutar, A. (2020). The impact of a self-avatar, hand collocation, and hand proximity on embodiment and stroop interference. In *IEEE Transactions on Visualization and Computer Graphics*, 26, 1964–1971. <https://doi.org/10.1109/TVCG.2020.2973061>
- Petersen, G. B., Petkakis, G., & Makransky, G. (2022). A study of how immersion and interactivity drive VR learning. *Computers & Education*, 179. <https://doi.org/10.1016/j.compedu.2021.104429>
- Pirker, J., Dengel, A., Holly, M., & Safikhani, S. (2020). Virtual reality in computer science education: A systematic review. In *Proceedings of the 26th ACM symposium on virtual reality software and*

- technology (Article 8, pp. 1–8). Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3385956.3418947>
- Plötzner, R., Berney, S., & Bétrancourt, M. (2021). When learning from animations is more successful than learning from static pictures: Learning the specifics of change. *Instructional Science*, 49, 497–514. <https://doi.org/10.1007/s11251-021-09541-w>
- Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality*, 20, 101–125.
- Reed, C. L., Grubb, J. D., & Steele, C. (2006). Hands up: Attentional prioritisation of space near the hand. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 166–177. <https://doi.org/10.1037/0096-1523.32.1.166>
- Reed, C. L., Betz, R., Garza, J. P., & Roberts, R. J., Jr. (2010). Grab it! Biased attention in functional hand and tool space. *Attention, Perception, & Psychophysics*, 72(1), 236–245. <https://doi.org/10.3758/APP.72.1.236>
- Remmele, M., Weiers, K., & Martens, A. (2015). Stereoscopic 3d's impact on constructing spatial hands-on representations. *Computers & Education*, 85, 74–83. <https://doi.org/10.1016/j.compedu.2015.02.008>
- Scharinger, C. (2018). Fixation-related EEG frequency band power analysis: A promising methodology for studying instructional design effects of multimedia learning material. *Frontline Learning Research*, 6, 57–71. <https://doi.org/10.14786/flr.v6i3.373>
- Scheiter, K., Brucker, B., & Ainsworth, S. (2020). “Now move like that fish”: Can enactment help learners come to understand dynamic motion presented in photographs and videos? *Computers & Education*, 155, 103934. <https://doi.org/10.1016/j.compedu.2020.103934>
- Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence*, 10, 266–281. <https://doi.org/10.1162/105474601300343603>
- Skulmowski, A. (2023a). Guidelines for choosing cognitive load measures in perceptually rich environments. *Mind, Brain, and Education*, 17, 20–28. <https://doi.org/10.1111/mbe.12342>
- Skulmowski, A. (2023b). Ethical issues of educational virtual reality. *Computers & Education: X Reality*, 2, 100023. <https://doi.org/10.1016/j.cexr.2023.100023>
- Skulmowski, A., & Rey, G. D. (2018). Embodied learning: Introducing a taxonomy based on bodily engagement and task integration. *Cognitive Research: Principles and Implications*, 3, 6. <https://doi.org/10.1186/s41235-018-0092-9>
- 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, 603–616. <https://doi.org/10.1162/pres.1997.6.6.603>
- Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE*, 5, 10564. <https://doi.org/10.1371/journal.pone.0010564>
- Smith, S. A. (2019). Virtual reality in episodic memory research: A review. *Psychonomic Bulletin & Review*, 26, 1213–1237. <https://doi.org/10.3758/s13423-019-01605-w>
- Stanney, K. M., & Hash, P. (1998). Locus of user-initiated control in virtual environments: Influences on cybersickness. *Presence*, 7(5), 447–459.
- Sun, R., Wu, Y. J., & Cai, Q. (2019). The effect of a virtual reality learning environment on learners' spatial ability. *Virtual Reality*, 23, 385–398. <https://doi.org/10.1007/s10055-018-0355-2>
- Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, 68, 1–16. <https://doi.org/10.1007/s11423-019-09701-3>
- Taylor, J. E., Gozli, D. G., Chan, D., Huffman, G., & Pratt, J. (2015). A touchy subject: Advancing the modulated visual pathways account of altered vision near the hand. *Translational Neuroscience*, 6, 1–7. <https://doi.org/10.1515/tnsci-2015-0001>
- Thorsteinsson, G. (2013). Developing an understanding of the pedagogy of using a virtual reality learning environment (VRLE) to support innovation education. In L. V. Shavinina (Ed.), *The Routledge international handbook of innovation education* (pp. 456–470). Abingdon: Routledge.
- Tseng, P., & Bridgeman, B. (2011). Improved change detection with nearby hands. *Experimental Brain Research*, 209, 257–269. <https://doi.org/10.1007/s00221-011-2544-z>
- Tseng, P., Bridgeman, B., & Juan, C.-H. (2012). Take the matter into your own hands: A brief review of the effect of nearby-hands on visual processing. *Vision Research*, 72, 74–77. <https://doi.org/10.1016/j.visres.2012.09.005>
- Tversky, B., Morrison, J. B., & Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, 57, 247–262. <https://doi.org/10.1006/ijhc.2002.1017>

- Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The mirror-neuron system and observational learning: Implications for the effectiveness of dynamic visualisations. *Educational Psychology Review*, 21, 21–30. <https://doi.org/10.1007/s10648-008-9094-3>
- Villena-Taranilla, R., Tirado-Olivares, S., Cózar-Gutiérrez, R., & González-Calero, J. A. (2022). Effects of virtual reality on learning outcomes in K-6 education: A meta-analysis. *Educational Research Review*, 35, 100434. <https://doi.org/10.1016/j.edurev.2022.100434>
- Wagler, A., & Hanus, M. D. (2018). Comparing virtual reality tourism to real-life experience: Effects of presence and engagement on attitude and enjoyment. *Communication Research Reports*, 35, 456–464. <https://doi.org/10.1080/08824096.2018.1525350>
- Weber, S.-C. (2016). *Touch and watch: The influence of hand proximity on knowledge acquisition with dynamic visualizations on multi-touch devices [Der Einfluss der Handnähe auf den Wissenserwerb mit dynamischen Visualisierungen auf Multi-Touch-Geräten]* [Unpublished bachelor thesis]. University of Tuebingen
- Wheeler, S. G., Engelbrecht, H., & Hoermann, S. (2021). Human factors research in immersive virtual reality firefighter training: A systematic review. *Frontiers in Virtual Reality*, 2, 671664. <https://doi.org/10.3389/frvir.2021.671664>
- Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, 51, 1991–2005. <https://doi.org/10.1111/bjet.13023>

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

Authors and Affiliations

Birgit Brucker¹  · Georg Pardi¹ · Fabienne Uehlin² · Laura Moosmann³ · Martin Lachmair⁴ · Marc Halfmann¹ · Peter Gerjets^{1,2}

✉ Birgit Brucker
b.brucker@iwm-tuebingen.de

¹ Leibniz-Institut Für Wissensmedien, Schleichstraße 6, 72076 Tübingen, Germany

² University of Tübingen, Schleichstraße 4, 72076 Tübingen, Germany

³ Furtwangen University, Kronenstraße 16, 78532 Tuttlingen, Germany

⁴ Baden-Württemberg Cooperative State University Villingen-Schwenningen, Friedrich-Ebert-Straße 30, 78054 Villingen-Schwenningen, Germany