



Burden or benefit? Effects of visuospatial imagery and fidelity on learning in immersive virtual reality

Bingyu Dong¹ · Yun Zhou¹ · Yueyao Qu¹ · Junjun Zhai¹ · Shu Zhao¹ · Jialin Yu¹ · Yihang Qin² · Tao Xu²

Received: 7 August 2025 / Accepted: 16 April 2026 / Published online: 10 May 2026
© The Author(s) 2026

Abstract

Immersive Virtual Reality (IVR) provides unique potential for facilitating STEM learning through realistic and interactive visualization of complex concepts. However, its effectiveness may be impacted by individual differences in visuospatial imagery (VSI), referring to learners' tendency or ability to visualize and mentally manipulate spatial information. Visual fidelity in IVR can serve as either a scaffold for mental imagery or a source of additional cognitive load, and its interaction with VSI remains unclear. To address this gap, we examined the effects of learning material fidelity (high vs. low) and self-perceived VSI (high vs. low) on learning fungal biology in IVR. This study adopted a 2×2 between-group design involving 120 college students (61 females, 59 males), and measured learning performance, cognitive load, motivation, learning experience, and emotions. Results revealed that VSI significantly moderated the effects of fidelity in IVR. Learners with low VSI benefited from high-fidelity learning materials, which compensated for their spatial processing limitations and improved retention, perceived ease of use, and positive emotions. In contrast, learners with high VSI performed similarly under both high- and low-fidelity conditions; however, they reported lower extraneous cognitive load, reduced fear of failure and negative emotions, and higher perceived usefulness, perceived ease of use, and satisfaction when using low-fidelity materials. These results extend the abilities-as-compensator hypothesis to IVR and self-reported VSI contexts and highlight the importance of adapting visual fidelity to learners' VSI when designing visualizations for STEM education.

Keywords Immersive virtual reality (IVR) · STEM · Fidelity · Visuospatial imagery (VSI) · Cognitive load · Learning experience

1 Introduction

Mastering STEM content typically requires students to process abstract and complex concepts that are often difficult to grasp through conventional media (Chng et al.

2023). Immersive Virtual Reality (IVR) technology, using head-mounted displays, headphones, and controllers, provides learners with a multisensory experience in a realistic digital space (Hartmann et al. 2023; Majewska and Vereen 2023). Research shows that IVR enhances engagement and

✉ Yun Zhou
zhouyun@snnu.edu.cn

Bingyu Dong
dongbingyu@snnu.edu.cn

Yueyao Qu
quyueyao@snnu.edu.cn

Junjun Zhai
zhaijunjun@snnu.edu.cn

Shu Zhao
zhaoshu@snnu.edu.cn

Jialin Yu
yujialin@snnu.edu.cn

Yihang Qin
qinyihang@mail.nwpu.edu.cn

Tao Xu
xutao@nwpu.edu.cn

¹ Faculty of Education, Shaanxi Normal University, South Chang'an Road 199, Yanta District, Xi'an 710062, Shaanxi Province, People's Republic of China

² School of Software, Northwestern Polytechnical University, 127 West Youyi Road, Beilin District, Xi'an 710072, Shaanxi Province, People's Republic of China

learning across diverse contexts. Studies have demonstrated its benefits for inquiry-based science learning (Yang et al. 2023), STEAM design and creativity (Izadinia 2023), and practical skills acquisition in science and engineering (Yang et al. 2024). Learners in immersive 360° or head-mounted display environments also show better conceptual understanding, attention, and emotional engagement than those using less immersive media (Rupp et al. 2019; Shadiev and Li 2023).

A key feature of IVR is its realistic visual representations, which have become an important design factor in educational IVR (Birbara and Pather 2021). However, the role of fidelity in learning remains debated: some studies suggest that highly realistic details increase cognitive load and hinder learning (e.g., Birbara and Pather 2021; Menendez et al. 2020), while recent findings indicate that high fidelity can facilitate effective mental model construction (e.g., Zhou et al. 2024). Although high fidelity enhances immersion, its impact on learning outcomes remains unclear, presenting challenges for educational IVR design.

Furthermore, the effects of realistic visualization in VR may vary across learners (Lin and Suh 2021), particularly depending on their visuospatial imagery (VSI), that is, their self-perceived or actual ability to mentally represent and manipulate spatial relationships (Hartmann et al. 2023; Hegarty et al. 2002). Low levels of VSI, or a lack of confidence in one's imagery ability, may hinder students from constructing mental representations of knowledge (Hartmann et al. 2023), especially in disciplines that rely heavily on spatial abilities (Safadel and White 2020). Therefore, understanding how VSI moderates the impact of visual fidelity is essential for optimizing STEM learning in IVR environments.

1.1 Cognitive scaffold or burden: effects of fidelity on learning and its theoretical background

Fidelity refers to the degree to which a simulation's appearance matches the real world's (Hays and Singer 1989). Historically, fidelity originated in the aviation field, as simulation entered the fields of medicine and education (Paige and Morin 2013). In VR, fidelity is an essential foundation for achieving realism: high fidelity often enhances realism by providing detailed and accurate representations that match real-world expectations. Realism in visualizations refers to how accurately a visual representation matches reality, involving factors like object shape, texture, and lighting (Skulmowski et al. 2022). These same factors are also central to the concept of fidelity, which similarly emphasizes the accuracy of visual elements (Kapralos et al. 2017). Although the terms fidelity and realism have different linguistic meanings, they are often used interchangeably

in the context of visualizations, with no significant philosophical distinction between them. Harrington provides a clear definition of different levels of fidelity (Harrington 2012), which is why this study adopts the term fidelity and follows the guidelines in his framework for design and development. Given their conceptual overlap, both terms are discussed in the following literature review.

Visual representations are widely employed as instructional tools in STEM contexts (Evagorou et al. 2015), and new technological advancements have led to increasingly realistic visualizations. However, the effects of such realism on learning remain inconclusive (Skulmowski 2022). Research on the effects of realistic visualization on learning has yielded divergent results, including positive (e.g., Menendez et al. 2022), negative (e.g., Scheiter et al. 2009), and neutral (e.g., Birbara and Pather 2021) outcomes. Several theoretical frameworks serve as the foundation for realistic visualization, with two of the most prominent being cognitive load theory and mental model theory. These theories provide opposing perspectives on the impact of realistic visualization. From the standpoint of cognitive load theory, highly realistic visuals may hinder learning by overloading cognitive resources. It posits that limited working memory constrains learners' ability to process information, thereby impacting performance (Linn and Petersen 1985; Sweller 1988). Realistic visualizations, by introducing more details and visual information, can increase extraneous cognitive load as a result. Consequently, excessive realism may overload perception and impair learning. For example, Skulmowski and Rey (2020) compared realistic and schematic 2D anatomical visualizations and found that realism increased extraneous cognitive load because of detailed imagery, although selectively applied realism improved retention. In contrast, in a virtual training study involving both cognitive and motor tasks, del Rocío Hernández-Castañón et al. (2023) compared learning under high and low levels of task abstraction and found that high-abstraction materials facilitated generalization of learning, whereas low-abstraction tasks promoted better transfer to similar tasks. Scheiter explored the effects of realistic and simplified dynamic learning resources on learning about cell mitosis, the results showed that simplified resources are better than realistic ones in testing (Scheiter et al. 2009). Similarly, Menendez et al. studied the impact of diagrammatic detail richness on adult learning of deformation knowledge, and the results indicated that simplified diagrams led to better transfer performance compared to detailed diagrams (Menendez et al. 2020).

However, recent studies suggest that realistic visualizations can be beneficial in some contexts. For instance, in an IVR study on object location memory, Huang and Klippel (2020) found that high visual realism afforded richer spatial

cues that supported spatial knowledge acquisition, though its effects were moderated by spatial ability. In a multi-representational computer-based learning study, adaptively increasing visual fidelity facilitated mental model construction by enhancing germane cognitive load (Joo et al. 2021). In diagram-based lessons on metamorphosis, perceptually rich visuals improved conceptual learning in younger learners but led to overgeneralization in older students (Menendez et al. 2022). In anatomy learning, realistic visualizations improved retention of specific details but increased extraneous cognitive load (Skulmowski and Rey 2020). More recently, in 3D animated geography lessons, high detail richness combined with low text load was found to improve immediate knowledge acquisition (Zhou et al. 2024). These studies indicate that increased visual fidelity or detail richness can facilitate mental model construction and conceptual learning, though their benefits depend on the learning context, modality, or learner characteristics. When the learning objective requires the engagement of mental imagery, incorporating detailed and realistic visuals can further enhance learning outcomes. While the added details in realistic visualizations may increase cognitive load, their positive effects on mental model construction often offset or even surpass these costs. Building on this line of evidence, researchers have recently proposed that realistic visualizations can function as cognitive scaffolds (Kaldaras et al. 2024), supporting learners in constructing mental models more effectively.

Taken together, the studies mentioned above primarily focus on traditional multimedia learning environments. While they do incorporate 3D learning resources, they differ significantly from IVR in key aspects. IVR, with its richer 3D visualizations compared to traditional multimedia, can serve as an effective scaffold for mental construction. However, it also places higher cognitive demands on learners and requires more attentional resources. As a result, learners may struggle to manage these demands within IVR. It is therefore not appropriate to directly transfer findings from multimedia research to IVR settings. Whether the mental scaffold function of IVR outweighs its cognitive burden, or vice versa, remains unclear. Taken together, further research is needed to explore how realistic visualizations impact on learning in IVR environments.

1.2 The impact of visuospatial imagery on learning: enhancer vs. compensator hypotheses

Although empirical evidence indicates that technology-enhanced environments, such as IVR, can enhance learners to visualize and learn (Ng et al. 2020), these environments often involve tasks that rely heavily on spatial processing, such as mentally positioning, transforming, and rotating

objects, assembling components, as well as depth perception and spatial memory. Visuospatial imagery, that is VSI, is generally regarded as a core component of the broader spatial ability construct (Hegarty et al. 2002). It has been defined as the ability to form vivid and clear spatial images when interacting with either real or mediated environments, conceptually overlapping with spatial intelligence (Hegarty 2010). Within multimedia and virtual environments, VSI is considered a key factor influencing both spatial understanding and the sense of presence. According to the process model of spatial presence proposed by Wirth et al. (2007), VSI plays a central cognitive role in constructing a spatial situation model and in the formation of spatial presence.

Previous research has adopted two main approaches to measure individual differences related to spatial processing: task-based performance tests (e.g., the Mental Rotation Test, Burte et al. 2019; Paper Folding Test, Cuendet et al. 2014) and self-report measures (e.g., the MEC Spatial Presence imagery subscale, Hartmann et al. 2023). Although these methods differ in operationalization, they are conceptually connected, as both tap into the cognitive processes involved in constructing and transforming mental representations of spatial relations. Prior evidence suggests that self-reported imagery tendencies correspond with performance-based spatial ability measures when both target similar forms of spatial processing (Blajenkova et al. 2006). Task-based measures assess objective spatial performance (Uttal and Cohen 2012; Wai et al. 2009), whereas self-reports capture subjective imagery tendency and confidence in using mental visualization strategies. For instance, the Vividness of Visual Imagery Questionnaire (VVIQ; Marks 1973) assesses the clarity and vividness of mental images (Iachini et al. 2019). Using the VVIQ, Alkhamash et al. (2025) found that learners with higher imagery vividness benefited most from VR-based anatomy learning, reporting stronger satisfaction, relevance, and presence. Iachini et al. (2019) showed that higher imagery vividness predicted stronger presence experiences in IVR, while Sunday et al. (2017) demonstrated that imagery clarity depends more on general imagery ability than on domain expertise.

Both perspectives are theoretically complementary: the former reflects learners' cognitive ability to process spatial information, whereas the latter captures their perceived readiness, confidence, and motivational disposition to engage in imagery-based reasoning during learning. In the context of IVR learning, self-perceived VSI is particularly relevant, as learners' subjective imagery vividness and sense of presence may directly influence their engagement and interaction with high- or low-fidelity visuals. Therefore, the present study adopts a self-perceived VSI perspective to examine how imagery tendencies interact with visual fidelity in impacting learning outcomes.

Visual-spatial ability affects the effectiveness of different instructional methods, which is explained by two hypotheses: the abilities-as-enhancer hypothesis and the abilities-as-compensator hypothesis. While the enhancer and compensator hypotheses were originally proposed to explain the moderating role of visual-spatial ability in learning (Mayer and Sims 1994; Brucker et al. 2014), their theoretical implications can be extended to self-perceived VSI given the close cognitive relationship between spatial ability and imagery tendency (Blajenkova et al. 2006). Both constructs involve the mental imagery, manipulation, and transformation of spatial information, differing primarily in whether they emphasize objective performance or self-perceived imagery tendency.

The abilities-as-enhancer hypothesis suggests that high visual-spatial ability enhances the effectiveness of the instruction, while low visual-spatial ability learners do not react sensitively to instruction (Brucker et al. 2014; Mayer and Sims 1994). In this view, individuals with high spatial ability can benefit more from 3D models with lower cognitive load (Huk 2006), as they can easily construct mental models (Mayer and Sims 1994), whereas 3D models may be burdensome for those with low spatial ability. Studies show that students with better mental rotation abilities benefit more from 3D visualizations in AR environments (Krüger et al. 2022), and 3D MR training is more effective than 2D MR training (Piri and Cagiltay 2023). Meta-analysis by Höffler also found that learners with high spatial ability perform better in visual learning (Höffler 2010).

The abilities-as-compensator hypothesis suggests that high visual-spatial abilities can compensate for the instruction (Brucker et al. 2014; Mayer and Sims 1994). Learners with lower spatial ability benefit more from 3D learning, as dynamic visualizations help them overcome difficulties in creating mental visualizations (Hays 1996). VR or 3D environments act as compensators, improving their learning performance (Teplá et al. 2022). Lee et al.'s study suggests that virtual learning environments have a more positive impact on these learners (Lee and Wong 2014). Additionally, learners with low spatial ability benefit more from spatial cues in CAVE environments (de Back et al. 2020).

Current research on visual-spatial ability has primarily examined its influence on learning with different visualization formats, such as 2D versus 3D materials (Brucker et al. 2014; Höffler 2010), yet relatively little attention has been paid to its interaction with visual fidelity in IVR. Despite these findings, most prior work has relied on objective measures of spatial ability, leaving the role of self-perceived VSI unclear. Given that self-perceived VSI reflects learners' perceived readiness, confidence, and motivation to engage in mental visualization, which may directly influence how they process high- or low-fidelity visuals in immersive

environments. Therefore, research is required to examine how self-reported VSI interacts with visual fidelity in affecting learning in IVR. Additionally, as IVR becomes more prevalent in STEM education, understanding this interaction will advance theoretical insights into enhancer and compensator effects in immersive learning.

1.3 The present study

To address this issue, this study examines the effects of learning material fidelity (high vs. low) and self-perceived VSI (high vs. low) on learning fungal biology in IVR and measures performance, cognitive load, attention, emotions, and learning experience. IVR has richer 3D visualizations compared to traditional multimedia, providing more detailed information for observation and the construction of mental models. While such visual richness may increase cognitive load, its potential benefits for supporting learners' understanding could outweigh these costs. This raises the possibility that the abilities-as-enhancer hypothesis may not be applicable in the context of IVR. Therefore, building on the abilities-as-compensator hypothesis, we propose an imagery-as-compensator account for IVR learning. When visuospatial demands are externalized through high-fidelity visualizations, learners with lower self-perceived VSI may benefit more, as the detailed external cues can compensate for their limited internal imagery. Conversely, learners with higher VSI may achieve better outcomes with lower-fidelity materials that reduce extraneous cognitive load.

Given this theoretical rationale, we formulate the following research questions and hypotheses to examine how self-perceived VSI interacts with the fidelity of IVR learning materials on STEM learning:

RQ1: How do VSI and fidelity influence learning performance in IVR?

H1: *Learners with low VSI will achieve higher retention (H1a) or transfer (H1b) in the high-fidelity condition compared to the low-fidelity condition. Learners with high VSI will achieve higher retention (H1c) or transfer (H1d) in the low-fidelity condition compared to the high-fidelity condition.*

RQ2: How do VSI and fidelity affect cognitive load in IVR?

H2: *Learners with low VSI will experience greater cognitive load in the high-fidelity condition compared to the low-fidelity condition (H2a). Learners with high VSI will*

experience greater cognitive load in the high-fidelity condition compared to the low-fidelity condition (H2b).

RQ3: How do VSI and fidelity influence attention in IVR?

H3: *Learners with low VSI will demonstrate higher levels of attention in the high-fidelity condition compared to the low-fidelity condition (H3a). Learners with high VSI will demonstrate higher levels of attention in the high-fidelity condition compared to the low-fidelity condition (H3b).*

RQ4: How do VSI and fidelity impact motivation in IVR?

H4: *Learners with low VSI will develop stronger motivation in the high-fidelity condition compared to the low-fidelity condition (H4a). Learners with high VSI will develop stronger motivation in the low-fidelity condition compared to the high-fidelity condition (H4b).*

RQ5: How do VSI and fidelity influence the learning experience in IVR?

H5: *Learners with low VSI will report a more positive learning experience in the high-fidelity condition compared to the low-fidelity condition (H5a). Learners with high VSI will report a more positive learning experience in the low-fidelity condition compared to the high-fidelity condition (H5b).*

RQ6: How do VSI and fidelity impact emotional responses in IVR?

H6: *Learners with low VSI will experience more positive emotions (H6a) or less negative emotions (H6b) in the high-fidelity condition compared to the low-fidelity condition. Learners with high VSI will experience more positive emotions (H6c) or less negative emotions (H6d) in the low-fidelity condition compared to the high-fidelity condition.*

Our predictions are grounded in the proposed imagery-as-compensator hypothesis (H1, H4, H5, and H6) and informed by cognitive load theory (H2 and H3). We also incorporate attention and motivation as key factors (Wu et al. 2022), since they may be influenced by varying levels of fidelity (H3 and H4). Considering the novelty of IVR technology and its interactive nature, we additionally assess satisfaction, perceived usefulness, and ease of use (H5). Emotions are closely linked to cognitive processes (Çağlar-Özhan and Altun 2020; Pelch 2018) with positive emotions enhancing motivation, focus, and performance,

while negative emotions increase cognitive load and reduce efficiency (Chang and Yang 2024). By examining emotional responses, we aim to understand the mechanisms through which fidelity and VSI affect learning. Therefore, we also measure emotions and proposed H6.

2 Methods

2.1 Participants and experimental design

This study used a 2×2 between-subjects design with learning material fidelity (high vs. low) and self-perceived VSI (high vs. low) as independent variables. Dependent variables included learning performance, cognitive load, attention, motivation, learning experience, and emotions. 120 participants (59 males, 61 females, aged 18–27, $M=21.60$, $SD=2.53$) were recruited from two universities. Participants were assigned to experimental conditions through a two-step procedure designed to balance VSI levels across fidelity groups. VSI was measured before the learning task. The first ten participants were randomly assigned to the high- or low-fidelity condition. For subsequent participants, the current sample median of accumulated VSI scores was used as a temporary threshold to classify them as having relatively higher or lower VSI. This rolling median fluctuated slightly between 37 and 38 during the early stage of data collection and stabilized at 38 in the later stage. Within each VSI stratum (higher vs. lower), participants were randomly assigned to the high- or low-fidelity condition to maintain balanced VSI distributions across groups. Table A1, and Figs A1 and A2 (in Appendix A) summarizes the descriptive statistics of the VSI scale, and Figure A3 shows the histogram of participants' VSI scores with a fitted normal curve, indicating an approximately unimodal, near-normal distribution. A post-hoc analysis confirmed a final sample median of 38, consistent with this allocation process (see Fig. A1 in Appendix A). The final group sizes were high visuospatial imagery with high fidelity (HVHF; $n=31$), high visuospatial imagery with low fidelity (HVLf; $n=31$), low visuospatial imagery with high fidelity (LVHF; $n=29$), and low visuospatial imagery with low fidelity (LVLf; $n=29$).

All groups learned the identical content, with fidelity as the only manipulated variable. Participants were screened for color vision deficiencies and VR motion sickness. Informed consent was obtained, and the study was approved by our institution's ethics committee. All participants were compensated. The required sample size was determined using G*Power 3.1.9.7 (Faul et al. 2007), based on an effect size of $\eta^2 = 0.09$ reported by Skulmowski (2022) and converted to $f=0.31$. With an effect

size of $f=0.31$, $\alpha=0.05$, and $\text{power}=0.80$, the required sample size for the four groups was estimated to be 118. The actual size meets this requirement.

2.2 IVR learning materials

The learning content was fungal biology, covering the concept of mushrooms, their classification, and five types of mushrooms in different growing environments in temperate forests. The five mushrooms and their respective environments were: *Amanita muscaria* in leaf litter, *Amanita regalis* under conifer trees, *Cortinarius fumosus* on decaying wood, *Mycena olivaceomarginata* in grasslands or meadows, and *Lepiota hemisphaerica* on dung piles.

The distinction between fidelity levels was based on Harrington's research on biology (Harrington 2012). We only manipulated the fidelity of the learning materials while the learning environment, tasks, and sequence were kept consistent across conditions. Accordingly, two VR programs corresponding to different fidelity levels were developed. Fidelity was operationalized through variations in texture resolution and geometric complexity. In the high-fidelity condition, high-resolution photographic textures (2048×2048 px) were applied to detailed 3D mushroom models, with a maximum Unity texture size of 8192×8192 px (8 K). In the low-fidelity condition, the same 3D models were simplified by reducing polygonal resolution (see Fig. 1A and B) and replacing photographic textures with flat color patterns (e.g., red and white for *Amanita muscaria*; see Fig. 1C and D), preserving only key morphological

features to achieve a cartoon-like style (see Fig. 1E and F). In terms of geometric complexity, we compared the number of underlying mesh triangles between the five sets of learning materials (e.g., *Amanita muscaria*). A Mann-Whitney U test showed that the high-fidelity models ($\text{Mdn}=4152.00$) contained significantly more triangles than the low-fidelity models ($\text{Mdn}=1312.00$), $U=2.000$, $z=-2.200$, $p=0.032$, $r=0.70$. Figure 1A, C and E demonstrate the low-fidelity learning materials, while Fig. 1B, D and F display the high-fidelity version.

To test the distinctiveness of perceived fidelity between the high- and low-fidelity VR programs used in this study, 23 university students (not involved in the formal experiment) were recruited to experience both versions in a counterbalanced design. Participants evaluated the perceived fidelity using a five-item, seven-point scale adapted from the Scene Realism dimension of the German VR Simulation Realism Scale (Poeschl and Doering 2013). The adapted scale demonstrated good reliability and validity ($\alpha=0.960$, $\text{KMO}=0.907$). A paired-samples t-test revealed that perceived fidelity was significantly higher for the high-fidelity condition ($M=6.05$, $SD=0.74$) than for the low-fidelity condition ($M=3.73$, $SD=1.61$), $t(22)=6.24$, $p<0.001$, Cohen's $d=1.30$. These results demonstrated a clear perceptual distinction between the two VR programs.

The IVR environment was created using Unity 3D and Blender, with an interactive UI. The scene featured a coniferous forest setting, with lighting to simulate daylight. It included multiple 3D models such as pine trees,

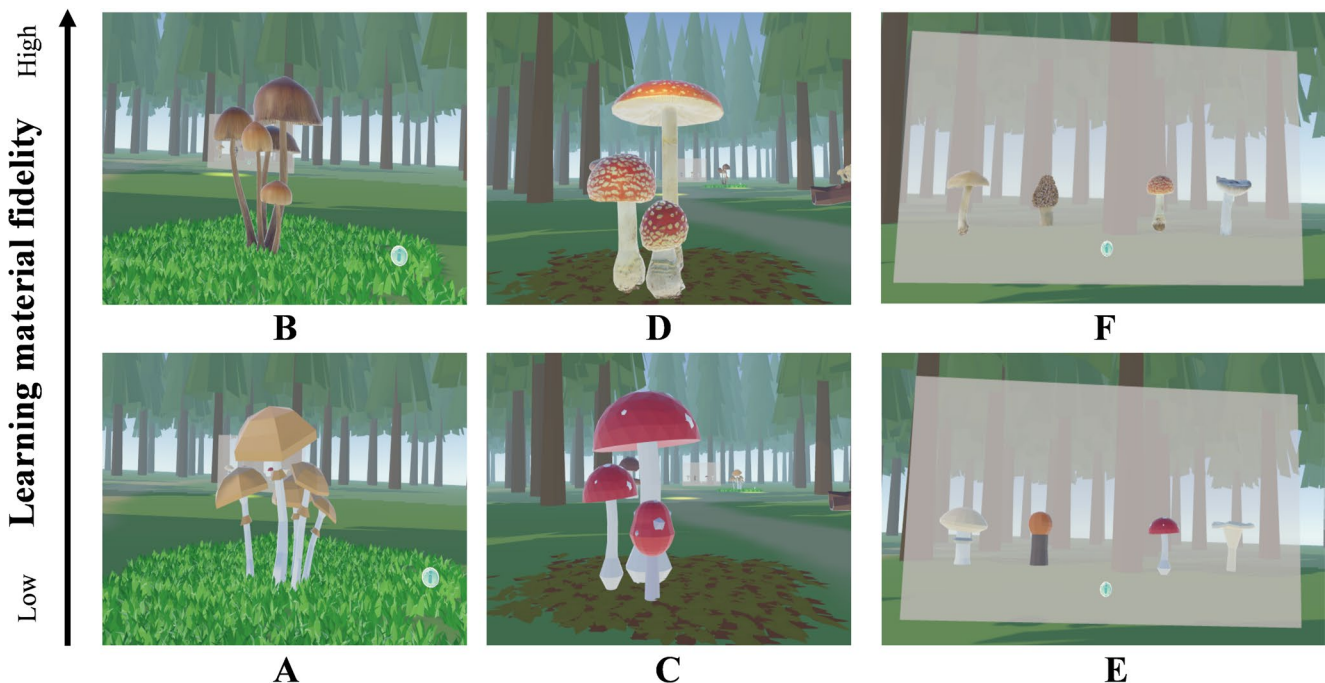


Fig. 1 The illustration of low-fidelity (A, C, and E) and high-fidelity (B, D, and F) learning materials

grass, mushrooms, interactive canvases, cameras, lights, and game objects. Upon entering the scene, participants followed the sequence of learning materials, each marked with a yellow ring on the ground for easy identification and navigation.

The IVR device used was the HTC VIVE Pro, which includes a head-mounted display, controllers, and trackers. The display has a combined resolution of 2880×1600 for both eyes, a 90 Hz refresh rate, built-in high-resolution headphones, and an interpupillary distance of 64 mm. It uses Steam VR Tracking for positional tracking, with interaction via VIVE controllers. The system was connected to an Alienware VR laptop running Windows 10 (8GB RAM, NVIDIA GeForce GTX 1060, Intel Core i7-7700). The program was controlled through the Steam VR plugin for IVR operations.

2.3 Measures

2.3.1 Visuospatial imagery test

Participants' visuospatial imagery was assessed using the eight-item MEC Spatial Presence Questionnaire (Hartmann et al. 2023), rated on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree; total score range 0–56, higher = stronger imagery). These scores were used during recruitment to classify participants' imagery levels based on a median-split approach (Lee and Wong 2014), but with a rolling median that was updated as new data accumulated, thereby supporting the stratified randomization procedure.

2.3.2 Learning performance

The learning performance test consisted of pre- and post-tests with 22 multiple-choice questions each, worth 1 point each for a total of 22 points. To prevent guessing, each question had five options: four answers and one "I do not know." The difficulty index (DFI) and differentiation index (DI) of the test were both satisfactory (DFI = 0.592, DI = 0.245). A DFI in the range of 0.5–0.6 is considered ideal, and a DI between 0.21 and 0.29 is acceptable (Baker 2001; Kumar et al. 2021). Both indices were calculated according to classical test theory, where the DFI reflects the proportion of correct responses and the DI indicates the extent to which an item differentiates between higher- and lower-performing students. Since the Cronbach's alpha was relatively low (0.503), a confirmatory factor analysis (CFA) was conducted to validate the learning performance (see Table A2 in Appendix A).

The test included retention and transfer parts, both consisting of multiple-choice questions. The retention test had 13 questions, both image- and text-based, such as "What color is the spore print of *Amanita regalis*?" The transfer test had 9 items, like "To promote wild mushroom growth in a city park, which measure should be taken based on *Lepiota hemisphaerica*'s growth preferences?"

2.3.3 Post-experiment subjective scales

Table 1 presents the information of all post-experiment subjective scales.

Cognitive load

The cognitive load test was adapted from (Skulmowski and Rey 2020), assessing intrinsic (Cronbach's alpha = 0.74) and extraneous cognitive load (Cronbach's alpha = 0.78).

Table 1 The descriptive information for all subjective scales

Variable	Subdimensions	Numbers of questions	Type	Reliability (Cronbach's alpha)	References
Cognitive load	Intrinsic	5	7-point Likert	0.74	Revised from (Skulmowski and Rey 2020)
	extraneous			0.78	
Attention		3		0.76	Revised from (Vorderer et al. 2004)
Motivation	Interest	2		0.82	Revised from (Rheinberg et al. 2001)
	Probability of success	2		0.67	
	Fear of failure	2		0.82	
	Challenge	2		0.69	
Learning experience	Perceived usefulness	3		0.93	(Ai-Lim Lee et al. 2010)
	Perceived ease of use	3		0.82	
	Satisfaction	5		0.93	
Emotions	Positive	10		0.82	(Watson et al. 1988)
	Negative	10			

Attention

The attention test, based on the Attention Allocation Questionnaire (Vorderer et al. 2004), initially had four dimensions. After translation and removing one redundant item, three items remained (Cronbach's $\alpha=0.76$).

Motivation

The motivation scale was adapted from Rheinberg et al. (2001) and included eight revised items across four dimensions: challenge (Cronbach's $\alpha = 0.69$), interest (0.82), probability of success (0.67), and fear of failure (0.82).

Learning experience

User experience was measured through perceived usefulness and perceived ease of use (Davis 1989). The scale, adapted from (Ai-Lim Lee et al. 2010), had Cronbach's $\alpha=0.93$ for perceived usefulness and Cronbach's $\alpha=0.82$ for ease of use. The satisfaction scale, with Cronbach's $\alpha=0.93$, measured participants' satisfaction with the learning environment, experience, instructional methods, and performance.

Emotions

Emotions were measured using the PANAS questionnaire (Watson et al. 1988), with ten words each for positive and negative emotions (e.g., "excited," "distressed"). Higher scores indicate stronger emotions. Cronbach's α was 0.85 for positive emotions and 0.93 for negative emotions.

2.4 Procedure

The experiment was conducted in a quiet room equipped with IVR devices. After providing informed consent, participants completed a VSI test and a pre-test on a laptop. They were then assigned to either the high- or low-fidelity group. Following a brief introduction, participants

engaged in a 15-minute learning session using a head-mounted display, after which they completed a knowledge test, subjective scales, and a short interview. The procedure is illustrated in Fig. 2.

2.5 Data analysis

Based on the results of the Shapiro-Wilk test and visual assessment of histograms and normal Q-Q plots, it was determined that all observations in the four conditions followed a normal distribution. Therefore, further analyses were conducted using the variance analysis. We used η^2 to measure effect size in the ANOVA, with effect size defined as small (0.01), medium (0.06), and large (0.14). Post hoc comparisons were conducted either when significant group differences were detected or when non-significant results were accompanied by a medium or larger effect size (Chen et al. 2018).

3 Results

3.1 Learning performance

3.1.1 Pretest

Table 2 presents the descriptive statistics for each dimension across the four conditions. Table 3 shows an overview of the main effect and interaction of independent variables.

ANOVA of pre-test data showed no significant differences ($F(3, 116)=1.23$, $p=0.304$, $\eta^2 = 0.03$), indicating comparable knowledge levels across the four groups.

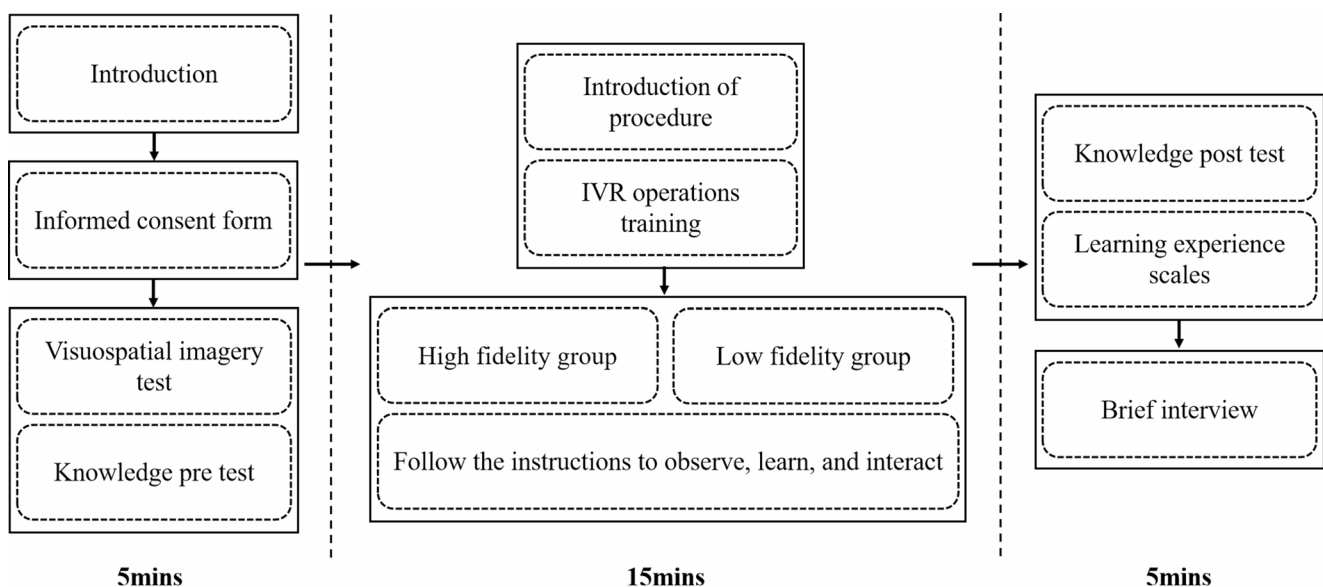
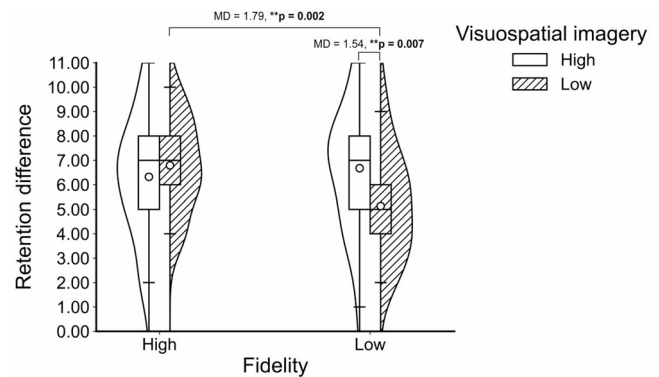


Fig. 2 Experimental procedure

Table 2 Descriptive statistics for each dimension under the four experimental conditions

Measures		HVHF (<i>n</i> = 31) M (SD)	HVLF (<i>n</i> = 31) M (SD)	LVHF (<i>n</i> = 29) M (SD)	LVLf (<i>n</i> = 29) M (SD)
Knowledge pretest		3.19 (2.66)	3.26 (2.57)	2.97 (2.32)	2.17 (2.21)
Knowledge posttest		13.13 (3.11)	14.23 (2.74)	13.83 (2.77)	11.90 (2.66)
Retention difference		6.19 (2.24)	6.68 (2.37)	6.93 (1.93)	5.14 (2.12)
Transfer difference		3.74 (2.05)	4.29 (2.19)	3.93 (1.81)	4.59 (1.90)
Cognitive load	Intrinsic	4.13 (1.35)	3.97 (1.41)	4.24 (1.39)	4.17 (1.18)
	Extraneous	3.72 (1.34)	2.79 (1.37)	3.15 (0.95)	3.30 (1.28)
Attention		5.74 (1.00)	6.01 (0.94)	5.41 (1.02)	5.15 (0.96)
Motivation	Interest	5.87 (0.89)	6.34 (0.79)	5.93 (0.98)	5.38 (1.73)
	Probability of success	5.42 (1.00)	5.95 (0.92)	5.50 (0.92)	5.50 (1.31)
	Fear of failure	3.24 (1.49)	2.53 (1.21)	3.28 (1.25)	3.72 (1.57)
	Challenge	5.36 (1.02)	5.90 (0.80)	5.40 (1.07)	5.10 (1.42)
Learning experience	Perceived usefulness	5.09 (1.29)	5.84 (1.02)	4.94 (1.20)	4.83 (1.46)
	Perceived ease of use	5.45 (0.93)	6.11 (0.71)	5.54 (0.94)	5.05 (1.11)
	Satisfaction	5.58 (0.91)	6.10 (0.70)	5.57 (1.01)	5.25 (1.35)
Emotions	Positive	5.03 (0.76)	5.14 (0.94)	5.04 (0.76)	4.50 (0.89)
	Negative	1.98 (1.03)	1.44 (0.45)	1.84 (0.99)	2.41 (1.32)

**Fig. 3** Differences in the retention difference across four conditions

3.1.2 Retention

A one-way ANOVA indicated significant differences in retention scores across the four groups ($F(3, 116) = 3.88$, $p = 0.011$, $\eta^2 = 0.09$). As shown in Fig. 3, post-hoc multiple comparisons revealed that the LVHF group outperformed the LVLf group (Mean difference (MD) = 1.79, $p = 0.002$), and the HVLF group outperformed the LVLf group (MD = 1.54, $p = 0.007$).

A two-way ANOVA revealed no significant main effects of VSI ($F(1,116) = 1.02$, $p = 0.315$, $\eta^2 = 0.01$) or fidelity ($F(1,116) = 2.71$, $p = 0.102$, $\eta^2 = 0.02$), but a significant interaction ($F(1,116) = 8.21$, $p = 0.005$, $\eta^2 = 0.07$; see Fig. 4). Each data point represents the mean value of one experimental group (HVHF, HVLF, LVHF, LVLf), and the connecting lines depict the interaction trend across fidelity levels for learners with high and low VSI. These mean-based plots are used to visualize the interaction between VSI and fidelity conditions.

Table 3 Main and interaction effects of independent variables

Measures		VSI		Fidelity		VSI × Fidelity	
		<i>p</i>	η^2	<i>p</i>	η^2	<i>p</i>	η^2
Retention difference		0.315	0.01	0.102	0.02	0.005**	0.07
Transfer difference		0.508	0.04	0.102	0.02	0.884	0.00
Overall performance		0.801	0.00	0.933	0.00	0.087	0.03
Cognitive load	Intrinsic	0.518	0.00	0.638	0.00	0.850	0.00
	Extraneous	0.901	0.00	0.088	0.03	0.019*	0.05
Attention		0.001**	0.09	0.990	0.00	0.140	0.02
Motivation	Interest	0.034*	0.04	0.842	0.00	0.017*	0.05
	Probability of success	0.333	0.01	0.166	0.02	0.166	0.02
	Fear of failure	0.017*	0.05	0.608	0.00	0.024*	0.04
	Challenge	0.061	0.03	0.525	0.00	0.038*	0.04
Learning experience	Perceived usefulness	0.013*	0.05	0.165	0.02	0.060	0.03
	Perceived ease of use	0.005**	0.07	0.634	0.00	0.001**	0.09
	Satisfaction	0.021*	0.05	0.592	0.00	0.026*	0.04
Emotions	Positive	0.046*	0.03	0.163	0.02	0.036*	0.04
	Negative	0.024*	0.04	0.941	0.00	0.003**	0.07

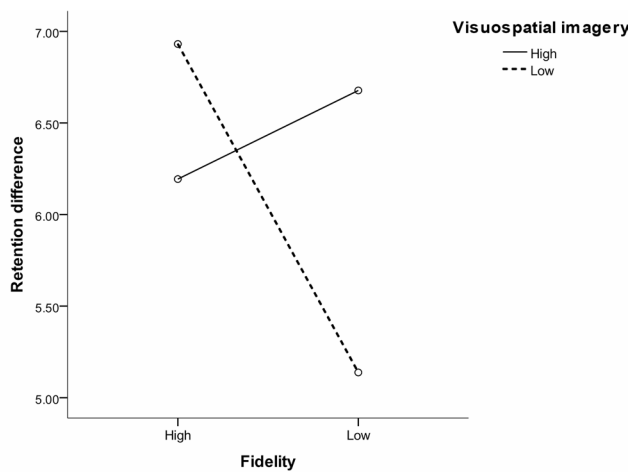


Fig. 4 Interaction between fidelity and visuospatial imagery in retention

3.1.3 Transfer

One-way ANOVA results indicated no significant differences in transfer scores across groups ($F(3, 116)=1.06$, $p=0.371$, $\eta^2 = 0.03$).

Two-way ANOVA revealed no significant main effects of VSI ($F(1, 116)=0.44$, $p=0.508$, $\eta^2 = 0.04$) or fidelity ($F(1, 116)=2.72$, $p=0.102$, $\eta^2 = 0.02$) on transfer difference, and no interaction effect ($F(1, 116)=0.02$, $p=0.884$, $\eta^2 = 0.00$).

3.1.4 Correlation between VSI and performance

A supplementary Pearson correlation analysis further confirmed this pattern: overall VSI was positively associated with learning performance ($r=0.205$, $p = 0.025$), but the relationship differed by fidelity condition: non-significant in the high-fidelity group ($r = -0.035$, $p = 0.793$) and strongly positive in the low-fidelity group ($r=0.454$, $p < 0.001$). These results reinforce the finding that high-fidelity IVR materials compensate for low imagery learners.

In summary, the results supported H1a but did not support H1b, H1c, or H1d.

3.2 Cognitive load

One-way ANOVA indicated no significant differences in overall cognitive load across the four groups ($F(3, 116)=2.11$, $p=0.103$, $\eta^2 = 0.05$).

The ANOVA results ($F(3, 116)=0.23$, $p=0.876$, $\eta^2 = 0.06$) and post-hoc comparisons showed no significant differences in intrinsic cognitive load across the four groups.

One-way ANOVA analysis indicated significant differences in extraneous cognitive load across the four groups

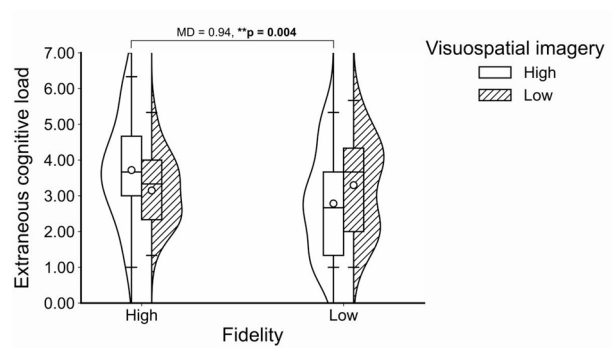


Fig. 5 Differences in extraneous cognitive load across four conditions

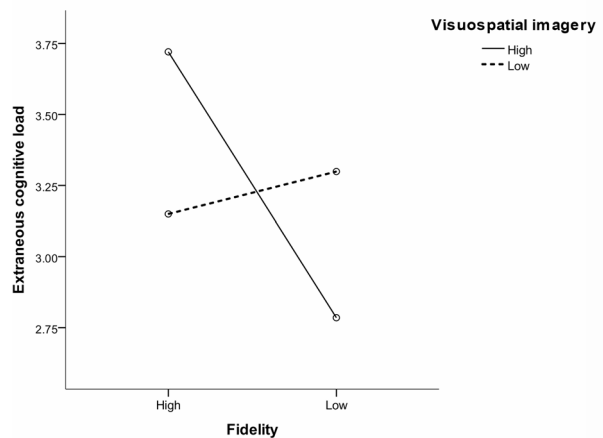


Fig. 6 Interaction between fidelity and visuospatial imagery in extraneous cognitive load

($F(3, 116)=2.97$, $p=0.035$, $\eta^2 = 0.07$). Post-hoc multiple comparisons revealed that HVHF group had significantly higher cognitive load than the HVLF group ($MD=0.94$, $p=0.004$; see Fig. 5).

Two-way ANOVA indicated no significant main effects of VSI or fidelity on intrinsic and extraneous cognitive load (see Table 3). However, a significant interaction effect was found on extraneous cognitive load ($F(1, 116)=5.65$, $p=0.019$, $\eta^2 = 0.05$; see Fig. 6).

The results rejected H2a but the results of extraneous cognitive load partially supported H2b.

3.3 Attention

One-way ANOVA showed significant differences in attention across the four groups ($F(3, 116)=4.40$, $p=0.006$, $\eta^2 = 0.10$). As shown in Fig. 7, post-hoc analysis revealed higher scores for the HVLF group compared to the LVLF group ($MD=0.86$, $p=0.001$) and the LVHF group ($MD=0.60$, $p=0.020$). The HVHF group also focused more than the LVLF group ($MD=0.59$, $p=0.021$).

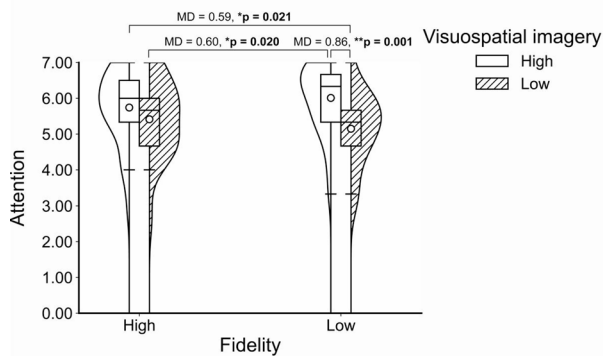


Fig. 7 Differences in attention across four conditions

Two-way ANOVA indicated a significant main effect of VSI ($F(1, 116)=10.60, p=0.001, \eta^2=0.09$) but not fidelity ($F(1, 116)=0.00, p=0.990, \eta^2=0.00$) on attention, with no interaction effect ($F(1, 116)=2.21, p=0.140, \eta^2=0.02$).

The results rejected H3a and b.

3.4 Motivation

One-way ANOVA results indicated no significant differences in overall motivation across the four groups ($F(3, 116)=0.76, p=0.518, \eta^2=0.02$). Two-way ANOVA indicated no significant main effects of VSI ($F(1, 116)=0.61, p=0.436, \eta^2=0.01$) or fidelity ($F(1, 116)=0.19, p=0.667, \eta^2=0.00$) on motivation, nor any significant interaction effect ($F(1, 116)=1.45, p=0.231, \eta^2=0.01$).

One-way ANOVA showed significant differences in interest across the four groups ($F(3, 116)=3.50, p=0.018, \eta^2=0.08$). Post-hoc analysis revealed that the HVLF group scored significantly higher than the LVLF group ($MD=0.96, p=0.002$; see Fig. 8).

One-way ANOVA indicated no significant differences in the probability of success across the four groups ($F(3, 116)=1.65, p=0.181, \eta^2=0.04$).

One-way ANOVA showed significant differences in fear of failure across the four groups ($F(3, 116)=3.79, p=0.012, \eta^2=0.09$). Post-hoc multiple comparisons revealed that the HVLF group scored significantly lower than the HVHF group ($MD=0.71, p=0.047$), LVHF group ($MD=0.74, p=0.041$), and LVLF groups ($MD=1.19, p=0.001$), as shown in Fig. 8.

One-way ANOVA results showed significant differences in challenge across the four groups ($F(3, 116)=2.84, p=0.041, \eta^2=0.07$). Post-hoc multiple comparisons revealed that the HVLF group scored significantly higher than the LVLF group ($MD=0.80, p=0.006$; see Fig. 8).

Two-way ANOVA revealed a significant main effect of VSI on interest ($F(1, 116)=4.58, p=0.034, \eta^2=0.04$) and fear of failure ($F(1, 116)=5.83, p=0.017, \eta^2=0.05$). The main effect of fidelity on motivation was not significant. Significant interactions were found on interest ($F(1, 116)=5.89, p=0.017, \eta^2=0.05$), fear of failure ($F(1, 116)=5.20, p=0.024, \eta^2=0.04$), and challenge ($F(1, 116)=4.42, p=0.038, \eta^2=0.04$).

Results did not support H4a but partially supported H4b.

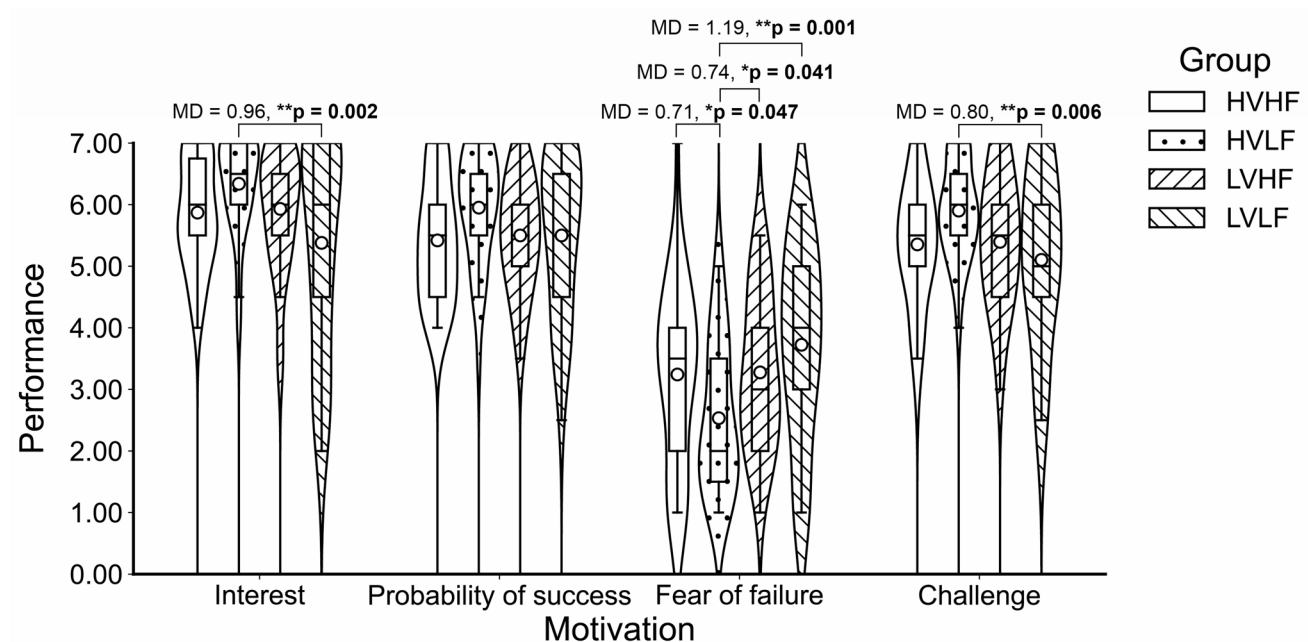


Fig. 8 Differences in motivation across four conditions

3.4.1 Learning experience

ANOVA analysis revealed significant differences in perceived usefulness across the groups ($F(3, 116)=4.06$, $p=0.009$, $\eta^2 = 0.10$). As shown in Fig. 9, post-hoc comparisons showed that the HVLF group scored significantly higher than the LVHF group ($MD=0.90$, $p=0.006$) and the LVLf group ($MD=1.01$, $p=0.002$). The HVHF group scored lower than HVLF group ($MD=0.75$, $p=0.019$).

For perceived ease of use, significant differences were found ($F(3, 116)=6.72$, $p=0.000$, $\eta^2 = 0.15$). As shown in Fig. 9, the HVLF group scored higher than the HVHF group ($MD=0.66$, $p=0.006$), the LVHF group ($MD=0.57$, $p=0.019$), and the LVLf group ($MD=1.06$, $p=0.000$). The LVHF group scored significantly higher than the LVLf group ($MD=0.49$, $p=0.045$).

One-way ANOVA also showed significant differences in satisfaction across the groups ($F(3, 116)=3.62$, $p=0.015$, $\eta^2 = 0.09$). As shown in Fig. 9, the HVLF group scored significantly higher than the HVHF group ($MD=0.52$, $p=0.047$), the LVHF group ($MD=0.53$, $p=0.045$), and the LVLf group ($MD=0.85$, $p=0.002$).

Two-way ANOVA revealed significant main effects of VSI on perceived usefulness ($F(1,116)=6.41$, $p=0.013$, $\eta^2 = 0.05$), perceived ease of use ($F(1,116)=8.26$, $p=0.005$, $\eta^2 = 0.07$), and satisfaction ($F(1,116)=5.43$, $p=0.021$, $\eta^2 = 0.05$). The main effects of fidelity on the learning experience and its dimensions were not significant. Interaction effects were found on perceived ease of use ($F(1,116)=11.55$,

$p=0.001$, $\eta^2 = 0.09$) and satisfaction ($F(1,116)=5.06$, $p=0.026$, $\eta^2 = 0.04$).

The results partially supported H5, with H5a partially supported by the results of perceived ease of use and H5b fully supported.

3.5 Emotions

A one-way ANOVA revealed significant differences across the four groups in both positive emotions ($F(3, 116)=3.45$, $p=0.019$, $\eta^2 = 0.08$) and negative emotions ($F(3, 116)=4.83$, $p=0.003$, $\eta^2 = 0.11$).

Post-hoc multiple comparisons showed that, in terms of positive emotions, the LVLf group scored significantly lower than HVLF group ($MD=0.64$, $p=0.004$), the HVHF group ($MD=0.53$, $p=0.017$), and the LVHF group ($MD=0.54$, $p=0.016$), as shown in Fig. 10.

For negative emotions, as shown in Fig. 10, the LVLf group scored significantly higher than the HVLF group ($MD=0.97$, $p=0.000$) and the LVHF group ($MD=0.57$, $p=0.032$). The HVHF group scored significantly higher than the HVLF group ($MD=0.54$, $p=0.035$).

A two-way ANOVA showed significant main effects of VSI on positive emotions ($F(1,116)=4.07$, $p=0.046$, $\eta^2 = 0.03$) and negative emotions ($F(1,116)=5.25$, $p=0.024$, $\eta^2 = 0.04$), while the main effect of fidelity was not significant on either positive emotions ($F(1, 116)=1.97$, $p=0.163$, $\eta^2 = 0.02$) or negative emotions ($F(1, 116)=0.01$, $p=0.941$, $\eta^2 = 0.00$). Additionally, significant interactions between

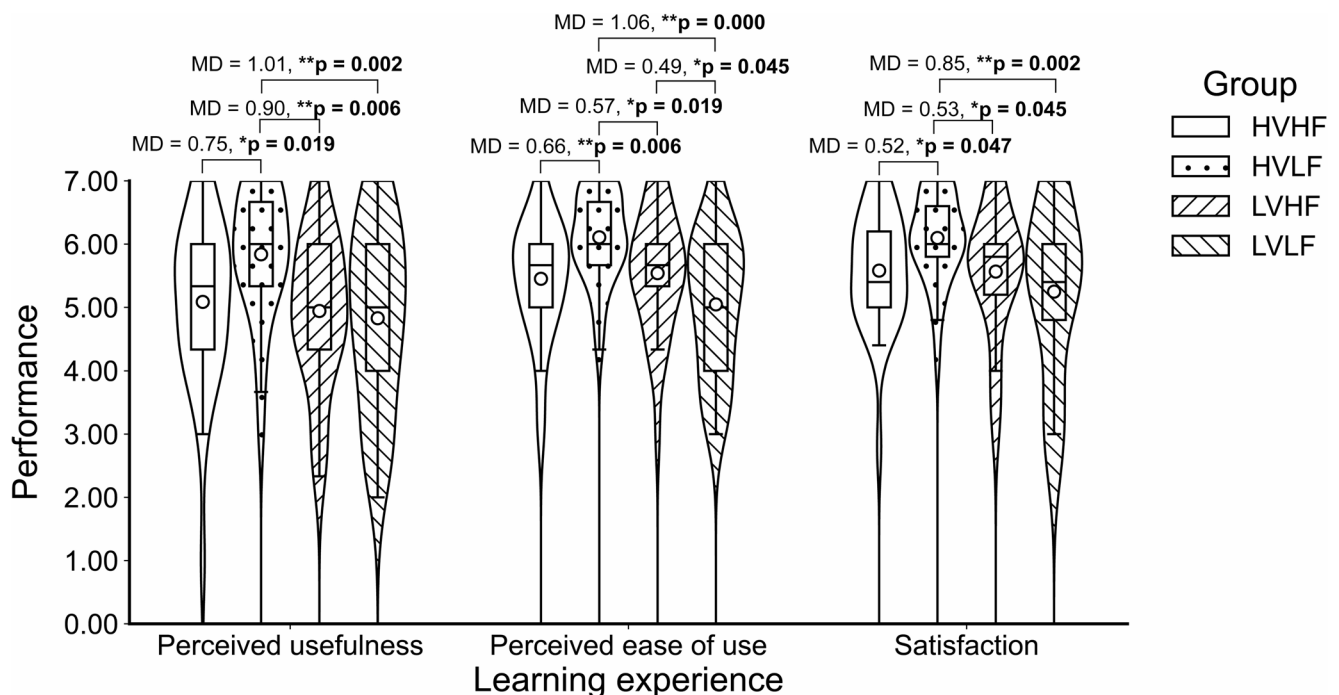
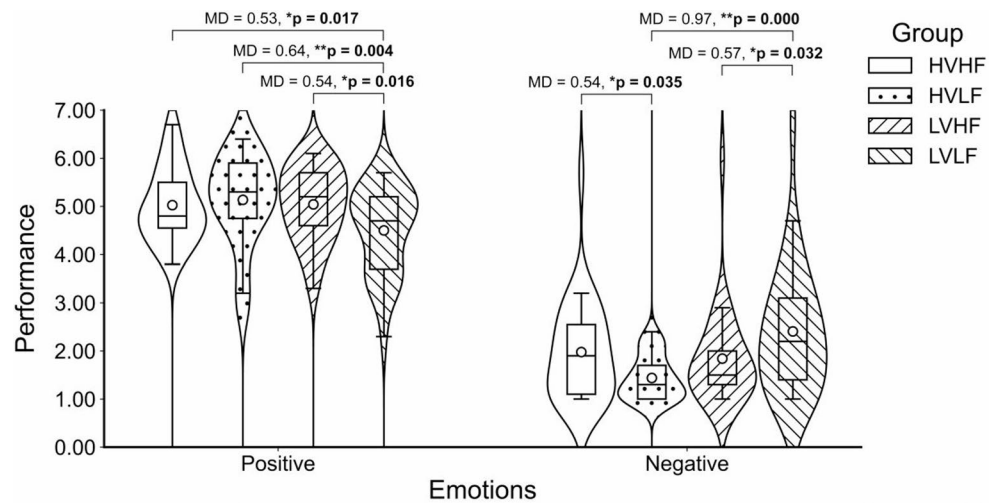


Fig. 9 Differences in learning experience across four conditions

Fig. 10 Differences in emotions across four conditions

VSI and fidelity were observed for both positive emotions ($F(1,116)=4.49$, $p=0.036$, $\eta^2 = 0.037$) and negative emotions ($F(1,116)=9.24$, $p=0.003$, $\eta^2 = 0.07$).

These results supported H6a, b, and d but did not support H6c.

4 Discussion

4.1 Empirical contributions

This study investigated how the fidelity of IVR learning materials interacts with self-perceived VSI to impact cognitive, motivational, and affective outcomes, revealing a consistent imagery-as-compensator pattern across dimensions. On learning performance, low-VSI learners achieved higher retention with high-fidelity visuals, consistent with the idea that detailed external cues can compensate for limited internal imagery and scaffold mental model construction (Hays 1996; Lee and Wong 2014; Zhou et al. 2024). In contrast, high-VSI learners performed similarly under both high- and low-fidelity conditions, indicating that they did not require high-fidelity visual details to support learning. This finding differs from prior empirical studies that did not consider individual differences in visuospatial ability, in which simplified visuals were thought to enhance learning by eliminating superfluous detail (Scheiter et al. 2009). For cognitive load, intrinsic load did not differ, but extraneous load showed a VSI and fidelity interaction: high fidelity reduced extraneous load for low-VSI learners by externalizing spatial information, yet increased extraneous load for high-VSI learners, who likely experienced redundant details as interference (Renkl and Scheiter 2017; Paas et al. 2003).

Parallel patterns appeared in attention, motivation, learning experience, and emotions. Attention was driven primarily by VSI rather than fidelity: high-VSI learners sustained

stronger visual engagement across conditions (Suh and Cho 2021), while low-VSI learners' attention remained constrained even with richer visuals. Motivationally, high-VSI learners reported lower fear of failure when learning with low-fidelity materials, whereas low-VSI learners were more motivated when aided by high-fidelity details in the learning materials (Chen and Wang 2018). These motivational differences mapped onto learning experience: high-VSI learners perceived low-fidelity IVR as more useful, easier to use, and more satisfying, aligning with work showing that display or interaction fidelity influences usability (McMahan et al. 2012). Affectively, high-VSI learners experienced fewer negative emotions with low-fidelity learning materials, while low-VSI learners benefited emotionally from high fidelity materials that reduced uncertainty and supported schema formation (Brockbank and Feldon 2024).

Compared with prior literature, our findings extend abilities-as-enhancer or compensator accounts, largely developed for objective spatial ability and 2D/3D multimedia, to self-perceived VSI in fully immersive VR. Whereas enhancer accounts typically predict that high-ability learners benefit most from complex and dynamic visuals (Huk 2006), our findings reveal a different pattern in IVR: high fidelity primarily supports low-VSI learners, who rely on external visual detail to compensate for limited internal imagery (Hays 1996; Lee and Wong 2014; Zhou et al. 2024). In contrast, high-VSI learners performed similarly under both high- and low-fidelity conditions, yet they experienced lower extraneous cognitive load and better learning experience when using low-fidelity materials. The results indicate that the effects of fidelity are not uniform across learners: learners with low visual-spatial ability perform better under high-fidelity conditions, whereas learners with high visual-spatial ability report more positive learning experiences under low-fidelity conditions, despite showing no significant differences in performance. This role reversal

helps explain the inconsistent findings regarding fidelity effects reported in prior studies.

Additionally, unlike performance-based ability measures, VSI captures learners' self-perceived readiness and confidence to engage in imagery-based reasoning, as well as the vividness and clarity of their mental visualizations. These aspects reflect learners' preparedness to process spatial information, their self-efficacy in mentally constructing and manipulating spatial representations, and to some extent, their willingness to engage in imagery-based processing during learning. This subjective dimension introduces a novel lens through which to understand learner variability in IVR, complementing traditional performance-based views of spatial ability.

4.2 Theoretical contributions

This study advances the theoretical understanding of how self-perceived VSI interacts with visual fidelity in IVR learning. While prior research on visualization and spatial ability has primarily been situated within multimedia learning and grounded in abilities-as-enhancer hypothesis (Brucker et al. 2014), our findings extend the abilities-as-compensator hypothesis to the IVR context, reframing it as an imagery-as-compensator framework. This framework emphasizes that the cognitive and affective benefits of high-fidelity IVR depend on learners' self-perceived imagery skills. For learners with low VSI, high-fidelity visualizations act as external cognitive scaffolds, compensating for limited internal imagery and improving learning. On the contrary, learners with high VSI performed similarly across fidelity levels, as their strong spatial abilities enable them to mentally construct and manipulate representations without relying on detailed external visuals.

From a broader theoretical perspective, this study integrates cognitive load theory and mental model theory to explain the mechanisms underlying these interactions. For low-VSI learners, high-fidelity learning materials alleviate the need for intensive internal visualization, thereby facilitating mental model formation and supporting comprehension of conceptual knowledge. High-VSI learners rely more on their internal imagery processes; while the added richness of high-fidelity materials increases extraneous cognitive load, it does not impair their schema construction or learning performance due to their strong spatial abilities. Therefore, the effects of fidelity in IVR depend on a balance between external scaffolding and internal cognitive economy.

Theoretically, this research broadens the understanding of fidelity beyond a universal design parameter by positioning it as a learner-dependent instructional scaffold. It complements findings from traditional multimedia environments

(e.g., 2D or screen-based learning) with evidence from fully immersive VR STEM contexts, where sensory realism fundamentally alters cognitive and affective processing. By linking self-perceived VSI to fidelity effects, this study provides a more nuanced model of learner-environment interaction and offers a theoretical foundation for adaptive fidelity design in immersive learning environments.

4.3 Practical contributions

Building on the evidence from this study, we recommend two design implications for IVR-based STEM learning environments. First, visual fidelity should be personalized to match learners' levels of VSI. For low-VSI learners, high-fidelity materials offer rich visual detail that serves as an external scaffold, compensating for their limited internal imagery and facilitating mental model construction. In contrast, high-VSI learners benefit from both fidelity levels; however, low-fidelity materials may reduce extraneous cognitive load and enhance their overall learning experience.

Second, for low-VSI learners, IVR environments should provide additional scaffolding, such as visual highlights, dynamic cues, or guided attention prompts, to direct focus toward relevant spatial relationships and reduce unnecessary cognitive load. These scaffolds can enhance comprehension while minimizing the risk of excessive extraneous cognitive load in highly detailed environments.

5 Limitations and future work

This study has several limitations. First, all assessments were immediate; future research could explore whether delayed assessments yield different results. Second, the study manipulated only the fidelity of learning materials within a standardized environment. Future work should investigate how varying fidelity levels of environments impact learning and experience. Third, to reflect the full-sample variability and preserve generalizability, the present study adopted a median-based grouping approach informed by and building on established practice (e.g., Lee and Wong 2014). Future studies drawing on larger and more diverse samples could examine whether more complex or clustered patterns of VSI emerge. Fourth, the VSI measure in this study was self-perceived rather than performance-based. While it reflects learners' perceived readiness, confidence, or motivation to engage in imagery-based reasoning, it does not directly assess their actual visuospatial skills. As a self-report measure, it may be influenced by biases related to factors such as age, gender, or self-efficacy. Future studies should therefore combine objective spatial ability tasks (e.g., Mental Rotation Test) with self-perceived measures to

validate findings and clarify the relationship between subjective VSI and objective spatial ability. Finally, this study focused on the acquisition of biological knowledge. STEM education, however, often involves procedural knowledge and skill-building, which may engage different cognitive processes compared to factual learning. Future studies could examine how fidelity affects learning on different types of knowledge learning and skills development.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10055-026-01380-9>.

Acknowledgements We appreciate all participants for their engagement in this research.

Author contributions Bingyu Dong and Yun Zhou contributed to the study conception and design. Material preparation, data collection and analysis were performed by Yueyao Qu and Junjun Zhai. Yihang Qin and Tao Xu assisted in refining the methodology. The first draft of the manuscript was written by Bingyu Dong and Yun Zhou. Shu Zhao and Jialin Yu contributed to reviewing and editing the manuscript. Also, Yun Zhou and Tao Xu provided the funding support for this research. All authors read and approved the final manuscript.

Funding This work was supported by the Research Projects of the National Natural Science Foundation of China under Grant [62377039]; the National Key R&D Program of China under Grant [2022YFC3303600]; CAAI-CANN Open Fund, developed on OpenI Community; Higher Education Research Fund of Northwestern Polytechnical University in 2025, No. GJJM202504.

Data availability The data are available from the corresponding author upon request.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval This research was approved by the Ethics Committee of the Faculty of Education at Shaanxi Normal University. All participants signed the informed consent agreement. The study followed the ethical guidelines of the institutional and national research committee, as well as WMA Declaration of Helsinki. All participants were informed of the voluntary nature of their participation and asked to provide written informed consent. They were assured of anonymity and informed about the purpose of the research, how their data would be used, and that no risks were associated with their involvement.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

References

- Alkhamash L, Saraco AN, Goes Schaurich C, Lo BSY, Krebs C (2025) The impact of an immersive virtual reality anatomy lab for informal science education. *Anat Sci Educ* 00:1–15. <https://doi.org/10.1002/ase.70087>
- Baker FB (2001) The basics of item response theory, 2nd edn. ERIC clearinghouse on assessment and evaluation
- Birbara NS, Pather N (2021) Real or not real: the impact of the physical fidelity of virtual learning resources on learning anatomy. *Anat Sci Educ* 14(6):774–787. <https://doi.org/10.1002/ase.2022>
- Blajenkova O, Kozhevnikov M, Motes MA (2006) Object-spatial imagery: a new self-report imagery questionnaire. *Appl Cogn Psychol* 20(2):239–263. <https://doi.org/10.1002/acp.1182>
- Brockbank RB, Feldon DF (2024) Cognitive reappraisal: the bridge between cognitive load and emotion. *Educ Sci* 14(8):8. <https://doi.org/10.3390/educsci14080870>
- Brucker B, Scheiter K, Gerjets P (2014) Learning with dynamic and static visualizations: realistic details only benefit learners with high visuospatial abilities. *Comput Hum Behav* 36:330–339. <https://doi.org/10.1016/j.chb.2014.03.077>
- Burte H, Gardony AL, Hutton A, Taylor HA (2019) Knowing when to fold'em: problem attributes and strategy differences in the paper folding test. *Pers Individ Differ* 146:171–181
- Çağlar-Özhan Ş, Altun A (2020) The cognitive-emotional processes and their implications for teacher education research. *J Psychol Educ Res* 28(2):55–74
- Chang C-C, Yang S-T (2024) Learners' positive and negative emotion, various cognitive processing, and cognitive effectiveness and efficiency in situated task-centered digital game-based learning with different scaffolds. *Interact Learn Environ* 32(9):5058–5077. <https://doi.org/10.1080/10494820.2023.2209600>
- Chen C-M, Wang J-Y (2018) Effects of online synchronous instruction with an attention monitoring and alarm mechanism on sustained attention and learning performance. *Interact Learn Environ*. <https://www.tandfonline.com/doi/abs/https://doi.org/10.1080/10494820.2017.1341938>
- Chen T, Xu M, Tu J, Wang H, Niu X (2018) Relationship between omnibus and post-hoc tests: an investigation of performance of the F test in ANOVA. *Shanghai Arch Psychiatr* 30(1):60–64. <https://doi.org/10.11919/j.issn.1002-0829.218014>
- Chng E, Tan AL, Tan SC (2023) Examining the use of emerging technologies in schools: a review of artificial intelligence and immersive technologies in STEM education. *J STEM Educ Res* 6(3):385–407. <https://doi.org/10.1007/s41979-023-00092-y>
- Cuendet S, Dehler-Zufferey J, Arn C, Bumbacher E, Dillenbourg P (2014) A study of carpenter apprentices' spatial skills. *Empir Res Vocat Educ Train* 6(1):3. <https://doi.org/10.1186/s40461-014-0003-3>
- Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q* 13(3):319–340
- de Back TT, Tinga AM, Nguyen P, Louwerse MM (2020) Benefits of immersive collaborative learning in CAVE-Based Virtual Reality. *Int J Educ Technol High Educ* 17(1):51
- del Rocío Hernández-Castañón V, Castillo-Ávila AA, Reyes-Meza V, Bianchi-Berthouze N, Morán AL, Orihuela-Espina F (2023) Effect of the level of task abstraction on the transfer of knowledge from virtual environments in cognitive and motor tasks. *Front Behav Neurosci* 17. <https://doi.org/10.3389/fnbeh.2023.1162744>

- Evagorou M, Erduran S, Mäntylä T (2015) The role of visual representations in scientific practices: from conceptual understanding and knowledge generation to 'seeing' how science works. *Int J STEM Educ* 2(1):11. <https://doi.org/10.1186/s40594-015-0024-x>
- Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39(2):175–191. <https://doi.org/10.3758/BF03193146>
- Harrington MCR (2012) The virtual trillion trail and the empirical effects of freedom and fidelity on discovery-based learning. *Virtual Real* 16(2):105–120. <https://doi.org/10.1007/s10055-011-0189-7>
- Hartmann C, Orli-Idrissi Y, Pflieger LCJ, Bannert M (2023) Imagine & immerse yourself: does visuospatial imagery moderate learning in virtual reality? *Comput Educ* 207:104909. <https://doi.org/10.1016/j.compedu.2023.104909>
- Hays TA (1996) Spatial abilities and the effects of computer animation on short-term and long-term comprehension. *J Educ Comput Res* 14(2):139–155. <https://doi.org/10.2190/60Y9-BQG9-80HX-UEML>
- Hays RT, Singer MJ (1989) Simulation fidelity as an organizing concept. In Hays RT, Singer MJ (eds), *Simulation fidelity in training system design* (pp. 47–75). Springer New York. https://doi.org/10.1007/978-1-4612-3564-4_3
- Hegarty M (2010) Chapter 7—components of spatial intelligence. In *Psychology of Learning and Motivation* (Vol. 52, pp. 265–297). Academic Press. [https://doi.org/10.1016/S0079-7421\(10\)52007-3](https://doi.org/10.1016/S0079-7421(10)52007-3)
- Hegarty M, Richardson AE, Montello DR, Lovelace K, Subbiah I (2002) Development of a self-report measure of environmental spatial ability. *Intelligence* 30(5):425–447. [https://doi.org/10.1016/S0160-2896\(02\)00116-2](https://doi.org/10.1016/S0160-2896(02)00116-2)
- Höfler TN (2010) Spatial ability: its influence on learning with visualizations—a meta-analytic review. *Educ Psychol Rev* 22(3):245–269. <https://doi.org/10.1007/s10648-010-9126-7>
- Huang J, Klippel A (2020) The effects of visual realism on spatial memory and exploration patterns in virtual reality. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology* (pp. 1–11). <https://doi.org/10.1145/3385956.3418945>
- Huk T (2006) Who benefits from learning with 3D models? The case of spatial ability. *J Comput Assist Learn* 22(6):392–404. <https://doi.org/10.1111/j.1365-2729.2006.00180.x>
- Iachini T, Maffei L, Masullo M, Senese VP, Rapuano M, Pascale A, Sorrentino F, Ruggiero G (2019) The experience of virtual reality: are individual differences in mental imagery associated with sense of presence? *Cogn Process* 20(3):291–298. <https://doi.org/10.1007/s10339-018-0897-y>
- Izadinia R (2023) I could feel a kind of keen air of excitement: using IVR to foster girls' confidence, interest, and engagement in STEAM. *J STEM Educ Res* 6(3):456–479. <https://doi.org/10.1007/s41979-023-00108-7>
- Joo H, Park J, Kim D (2021) Visual representation fidelity and self-explanation prompts in multi-representational adaptive learning. *J Comput Assist Learn* 37(4):1091–1106. <https://doi.org/10.1111/jcal.12548>
- Kaldaras L, Wang KD, Nardo JE, Price A, Perkins K, Wieman C, Salehi S (2024) Employing technology-enhanced feedback and scaffolding to support the development of deep science understanding using computer simulations. *Int J STEM Educ* 11(1):30. <https://doi.org/10.1186/s40594-024-00490-7>
- Kapralos B, Moussa F, Collins K, Dubrowski A (2017) Fidelity and multimodal interactions. In Wouters P, van Oostendorp H (eds.), *Instructional techniques to facilitate learning and motivation of serious games* (pp. 79–101). Springer International Publishing. https://doi.org/10.1007/978-3-319-39298-1_5
- Krüger JM, Palzer K, Bodemer D (2022) Learning with augmented reality: impact of dimensionality and spatial abilities. *Comput Educ Open* 3:100065. <https://doi.org/10.1016/j.caeo.2021.100065>
- Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivas V (2021) Item analysis of multiple choice questions: a quality assurance test for an assessment tool. *Med J Armed Forces India* 77:S85–S89. <https://doi.org/10.1016/j.mjafi.2020.11.007>
- Lee EA-L, Wong KW (2014) Learning with desktop virtual reality: low spatial ability learners are more positively affected. *Comput Educ* 79:49–58. <https://doi.org/10.1016/j.compedu.2014.07.010>
- Lee EA, Wong KW, Fung CC (2010) How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Comput Educ* 55(4):1424–1442. <https://doi.org/10.1016/j.compedu.2010.06.006>
- Lin Y, Suh A (2021) The role of spatial ability in learning with virtual reality: a literature review. <http://hdl.handle.net/10125/70622>
- Linn MC, Petersen AC (1985) Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Dev* 56(6):1479–1498. <https://doi.org/10.2307/1130467>
- Majewska AA, Vereen E (2023) Using immersive virtual reality in an online biology course. *J STEM Educ Res* 6(3):480–495. <https://doi.org/10.1007/s41979-023-00095-9>
- Marks DF (1973) Visual imagery differences in the recall of pictures. *Br J Psychol* 64(1):17–24. <https://doi.org/10.1111/j.2044-8295.1973.tb01322.x>
- Mayer RE, Sims VK (1994) For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *J Educ Psychol* 86(3):389–401. <https://doi.org/10.1037/0022-0663.86.3.389>
- McMahan RP, Bowman DA, Zielinski DJ, Brady RB (2012) Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE Transactions on Visualization and Computer Graphics*. 18(4):626–633. <https://doi.org/10.1109/TVCG.2012.43>
- Menendez D, Rosengren KS, Alibali MW (2020) Do details bug you? Effects of perceptual richness in learning about biological change. *Appl Cogn Psychol* 34(5):1101–1117. <https://doi.org/10.1002/acp.3698>
- Menendez D, Rosengren KS, Alibali MW (2022) Detailed bugs or bugging details? The influence of perceptual richness across elementary school years. *J Exp Child Psychol* 213:105269. <https://doi.org/10.1016/j.jecp.2021.105269>
- Ng O-L, Shi L, Ting F (2020) Exploring differences in primary students' geometry learning outcomes in two technology-enhanced environments: dynamic geometry and 3D printing. *Int J STEM Educ* 7(1):50. <https://doi.org/10.1186/s40594-020-00244-1>
- Paas F, Renkl A, Sweller J (2003) Cognitive load theory and instructional design: recent developments. *Educ Psychol* 38(1):1–4. https://doi.org/10.1207/S15326985EP3801_1
- Paige JB, Morin KH (2013) Simulation fidelity and cueing: a systematic review of the literature. *Clin Simul Nurs* 9(11):e481–e489. <https://doi.org/10.1016/j.ecns.2013.01.001>
- Pelch M (2018) Gendered differences in academic emotions and their implications for student success in STEM. *Int J STEM Educ* 5(1):33. <https://doi.org/10.1186/s40594-018-0130-7>
- Piri Z, Cagiltay K (2023) Can 3-dimensional visualization enhance mental rotation (MR) ability? A systematic review. *Int J Hum Comput Interact*. <https://doi.org/10.1080/10447318.2023.2196161>
- Poeschl S, Doering N (2013) The German VR simulation realism scale—psychometric construction for virtual reality applications with virtual humans. *Annual Rev Cybertherapy Telemed* 2013:33–37
- Renkl A, Scheiter K (2017) Studying visual displays: how to instructionally support learning. *Educ Psychol Rev* 29(3):599–621. <https://doi.org/10.1007/s10648-015-9340-4>

- Rheinberg F, Vollmeyer R, Burns B (2001) QCM: a questionnaire to assess current motivation in learning situations. *Diagnostica* 47(2):57–66
- Rupp MA, Odette KL, Kozachuk J, Michaelis JR, Smither JA, McConnell DS (2019) Investigating learning outcomes and subjective experiences in 360-degree videos. *Comput Educ* 128:256–268. <https://doi.org/10.1016/j.compedu.2018.09.015>
- Safadel P, White D (2020) Effectiveness of computer-generated virtual reality (VR) in learning and teaching environments with spatial frameworks. *Appl Sci* 10(16). <https://doi.org/10.3390/app10165438>
- Scheiter K, Gerjets P, Huk T, Imhof B, Kammerer Y (2009) The effects of realism in learning with dynamic visualizations. *Learn Instr* 19(6):481–494. <https://doi.org/10.1016/j.learninstruc.2008.08.001>
- Shadiev R, Li D (2023) A review study on eye-tracking technology usage in immersive virtual reality learning environments. *Comput Educ* 196:104681. <https://doi.org/10.1016/j.compedu.2022.104681>
- Skulmowski A (2022) Realistic visualizations can aid transfer performance: do distinctive shapes and descriptive labels contribute towards learning? *J Comput Assist Learn* 38(3):681–691. <https://doi.org/10.1111/jcal.12640>
- Skulmowski A, Rey GD (2020) The realism paradox: Realism can act as a form of signaling despite being associated with cognitive load. *Hum Behav Emerg Technol* 2(3):251–258. <https://doi.org/10.1002/hbe2.190>
- Skulmowski A, Nebel S, Remmele M, Rey GD (2022) Is a preference for realism really naive after all? A cognitive model of learning with realistic visualizations. *Educ Psychol Rev* 34(2):649–675. <https://doi.org/10.1007/s10648-021-09638-1>
- Suh J, Cho JY (2021) A triangular relationship of visual attention, spatial ability, and creative performance in spatial design: an exploratory case study. *J Interior Des* 46(3):11–27. <https://doi.org/10.1111/joid.12194>
- Sunday M, McGugin RW, Gauthier I (2017) Domain-specific reports of visual imagery vividness are not related to perceptual expertise. *Behav Res Methods* 49(2):733–738. <https://doi.org/10.3758/s13428-016-0730-4>
- Sweller J (1988) Cognitive load during problem solving: effects on learning. *Cogn Sci* 12(2):257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- Teplá M, Teplý P, Šmejkal P (2022) Influence of 3D models and animations on students in natural subjects. *Int J STEM Educ* 9(1):65. <https://doi.org/10.1186/s40594-022-00382-8>
- Uttal DH, Cohen CA (2012) Spatial thinking and STEM education. In *The Psychology of Learning and Motivation* (pp. 147–181). <https://doi.org/10.1016/b978-0-12-394293-7.00004-2>
- Vorderer P, Wirth W, Gouveia F, Biocca F, Saari T, Jäncke L, Böcking S, Schramm H, Gysbers A, Hartmann T, Klimmt C, Laarni J, Ravaja N, Sacau A, Baumgartner T, Jäncke P (2004) MEC spatial presence questionnaire (MEC-SPQ, English and German version): short documentation and instructions for application. In Report to the European Community, Project Presence: MEC (IST-2001-37661). <https://doi.org/10.13140/RG.2.2.26232.42249>
- Wai J, Lubinski D, Benbow CP (2009) Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J Educ Psychol* 101(4):817–835. <https://doi.org/10.1037/a0016127>
- Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol* 54(6):1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Wirth W, Hartmann T, Böcking S, Vorderer P, Klimmt C, Schramm H, Saari T, Laarni J, Ravaja N, Gouveia FR, Biocca F, Sacau A, Jäncke L, Baumgartner T, Jäncke P (2007) A process model of the formation of spatial presence experiences. *Media Psychol* 9(3):493–525. <https://doi.org/10.1080/15213260701283079>
- Wu C-H, Liu C-H, Huang Y-M (2022) The exploration of continuous learning intention in STEAM education through attitude, motivation, and cognitive load. *Int J STEM Educ* 9(1):35. <https://doi.org/10.1186/s40594-022-00346-y>
- Yang H, Cai M, Diao Y, Liu R, Liu L, Xiang Q (2023) How does interactive virtual reality enhance learning outcomes via emotional experiences? A structural equation modeling approach. *Front Psychol* 13. <https://doi.org/10.3389/fpsyg.2022.1081372>
- Yang C, Zhang J, Hu Y, Yang X, Chen M, Shan M, Li L (2024) The impact of virtual reality on practical skills for students in science and engineering education: a meta-analysis. *Int J STEM Educ* 11(1):28. <https://doi.org/10.1186/s40594-024-00487-2>
- Zhou Y, Yi F, Dong B, Zhang G, Zhang Y, Xu T (2024) Mental imagery scaffolding: the effects of detail richness and text load on geography learning. *Educ Inform Technol*. <https://doi.org/10.1007/s10639-024-12540-2>

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