

Virtual reality for student learning: Understanding individual differences

Ping Li^{1*}, Jennifer Legault², Alexander Klippel³, Jiayan Zhao³

Virtual reality (VR) is emerging as a rapidly developing technology that holds significant promises to impact student learning. In this review, we focus on the features of this technology regarding levels of immersion and interaction and individual differences in cognitive characteristics of VR learners. We attempt to parse the specific technological features that enable effective learning and examine how students mentally process these features. While VR helps to create situated learning conditions, its theoretical significance lies in its ability to provide perception-action enabled experiences to the learner, and it is these experiences that lead to positive behavioural and brain outcomes compared to traditional methods of learning. Our discussion highlights the understanding of VR learning with respect to individual differences, especially in spatial abilities of the learner, and how variability in spatial abilities might impact both spatial learning and language learning.

Keywords: virtual reality, embodied cognition, spatial learning, language learning, neuroplasticity, individual differences

INTRODUCTION

Virtual reality (VR) is in for education. Tech companies – big or small – have invested billions of dollars in recent years on platforms and tools that deploy virtual reality, augmented reality or mixed reality technologies, not just for gaming but also for student learning. As one example, *Google Expeditions* has developed over 100 VR trips for site visits (e.g. museums, historical relics, parks), used by students and teachers for learning science, arts and history. Other companies have also developed tools and software for education by making VR more portable, flexible and affordable. In response to VR's potential for transforming student learning, the US Department of Education set out the EdSim Challenge, an initiative “calling upon the virtual reality, video game developer and educational technology communities to submit concepts for immersive simulations that will prepare students for a globally competitive workforce and spur an ecosystem of virtual and augmented reality technology in education” (1).

VR, as the name suggests, is “reality simulated virtually” – a digitally presented real-world-like environment for users to see, play and socially interact. The simulated virtual environment contains dynamic images of objects and scenes, within which the VR user can move around and/or perform actions. The simulation is usually achieved via presentations or projections on displays of desktop computers, tablets, smartphones, head-mounted devices, large open walls or screens of TVs and amphitheatres. The VR industry manufactures many different types of VR products which vary significantly in the delivery of the quality of the simulated reality (see further discussion below). Such products also include tools for interactive perceptual experiences through digital image enhancement, i.e. “augmented reality” and a blend of real-world objects and actions with digitally presented virtual realities, i.e. “mixed reality”.

With the enormous public interest in VR, the question arises as to how we can take full advantage of the features of this technology for student learning. Despite the excitement about VR's application in various domains, significant gaps exist between what the industry develops and what the academics know about how students learn. VR has been shown to help improve people's lives in many domains, including aiding in sensorimotor training

and rehabilitation (2, 3), providing effective spatial navigation for patients with amnesia (4) and producing effective instructions for science (5) and second language (L2) learning (6, 7). However, our knowledge of the individual user/learner's variable experience in VR and how that impacts learning remains largely unknown. The current paper addresses the issue of VR efficacy for student learning from the individual differences' perspective. We attempt to provide an overview and synthesis of how the environment-learner interaction may produce positive impacts and changes on the mind and brain of the student learner. Our focus will be on separating technological features of VR from human experiences and, in doing so, identifying the role of learner abilities (e.g. cognitive, linguistic, spatial abilities) and how individual differences in these abilities impact VR learning efficacy.

FEATURES OF VR AND SITUATED LEARNING

While VR has offered great promises for enhancing education, determining the specific features of VR that lead to learning success is a matter of ongoing investigation (see (8) for a discussion). Here it is important for us to separate what the VR technology provides and what the human user/learner experiences in VR environments. We first discuss the technological side and identify important features of VR.

Immersion and interaction in VR

VR relies on advances in digital technology to simulate real-world environments, although its applications can go beyond real-world realities, e.g. simulating Shakespeare's stages or future forestation (9). The term “VR” has been used in the literature to refer to a wide range of virtual simulations, from the more basic or primitive three-dimensional (3D) dynamic images or videos to the fully immersive VR (iVR) experiences enabled by the use of gears such as head-mounted displays, haptic gloves and treadmills. (For ease of reference, we use the term “VR” to refer to different varieties of virtual reality, including augmented and mixed reality.)

First, VR can provide different degrees of *immersion* – a technical term used to describe the perception of being physically present in a non-physical world. In the current review, we posit that the level of immersion is determined by three main factors: 1) representational fidelity – how realistic the VR environments are designed to represent real objects and scenes, with regard to, for example, the quality of 3D models in terms of number of pixels per image and image details such as texture, colour, size and shape; 2) display fidelity – how the 3D images are presented, for example,

¹Department of Chinese and Bilingual Studies, Faculty of Humanities, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China. ²Department of Psychology, The Pennsylvania State University, University Park, 16802, PA, USA. ³Department of Geography, The Pennsylvania State University, University Park, 16802, PA, USA.

*Corresponding author. Email: pi2li@polyu.edu.hk (P.L.)

on high-definition displays/screens; 3) field of view (FOV) – how a convincing sense of space is created, for example, by using a confined viewing space (e.g. a 12-inch wide computer monitor) or a wide angle of view from a head-mounted display. These dimensions jointly affect the degree of immersion in VR; in particular, increased fidelity and FOV enhance the sense of presence such that users perceive themselves physically present in a non-physical world, even to the extent of losing awareness of the fact that they are actually in an artificial world.

Second, current VR technologies provide different VR platforms or interfaces, particularly in connection with the aforementioned features – display fidelity and FOV. At the less immersive end of the spectrum are desktop-based virtual environments, commonly found in gaming software. When we discuss desktop-based virtual environments for gaming and education, we refer to them as “desktop VEs” as opposed to “desktop VR” in order to maximally distinguish them from iVR and to avoid confusion. iVR environments today include the use of iVR devices such as head-mounted displays, through which the user can see a spatial layout that is dynamically changing in accord with a first-person’s perspective. In other words, turning one’s head will lead to a different view of objects in the scene, as in real-life situations. In contrast to presentations on a desktop computer or a TV screen, iVR conveys a comparatively greater feeling of “being there” or “spatial presence” (10). Spatial presence is the result of an increased level of immersion, achieved from 3D movies and IMAX theatres but less so from desktop platforms.

Third, recent technological developments have placed more emphasis on action performance in virtual environments. Actions include, at the simplest level, user visual interactions with objects (e.g. looking at an animated animal in a zoo), direct manipulation of objects through hand movements (e.g. picking up a virtual cup in a kitchen) or other bodily movements and locomotion (e.g. navigating a virtual town). More realistic actions involve simulated social interactions, e.g. dialogues with a virtual agent in iVR or multi-player user interactions within the same iVR environment (enabled by new technological development such as *HTC Vive-sync*). Finally, at a more complex level, actions can be performed in mixed reality (e.g. using *Microsoft HoloLens*), i.e. in both VR and real environments; for example, bouncing a virtual ball off a real table.

The significance of interactions (visual, manual or bodily) in iVR has been recognised by many scholars (see (11, 12) for reviews). The effects of action performance in manipulating objects or navigating in virtual environments are important for student learning and will be further discussed. Although IMAX theatres can provide a fairly high level of immersion through high display fidelity of 3D images and panoramic vision, movie contents are only passively received by the viewer, that is, without the viewer’s active interactions with the scenes. Desktop-based gaming VEs (e.g. *Second Life*), on the other hand, may enable some interactions but these interactions need to be mediated through an avatar which is controlled via a computer mouse or a remote device and therefore do not provide a high level of immersion. In contrast, iVR allows “whole-body” simulations in a first-person’s perspective which in turn allow highly realistic interactions, maximising the virtual experience. Users therefore feel highly immersed in the experience and may even have a temporary suspension of belief (“disbelief”) that he or she is “inside” a digital environment (11). This “disbelief” is the basis for some intense physical and psychological responses within the environment, a feature that the gaming industry wants

to achieve: the greater the disbelief (e.g. fighting against a virtual monster), the more effective the virtual experience becomes.

Desktop-based virtual environments (VEs) versus immersive VR (iVR)

Due to the high prevalence of desktop-based virtual environments (VEs henceforth for consistency), it is necessary for us to briefly compare desktop-based VEs with iVR in order to identify the important features for virtual learning.

First, for desktop VEs, immersion is usually mediated through the use of a personalised avatar, with the users imagining or contextualising themselves in the virtual space. In contrast, contextualisation is not required in iVR due to the use of whole-body simulations in first-person’s perspective. Second, the features of objects and scenes on a desktop are presented on a two-dimensional (2D) screen with limited FOV (e.g. the limited size of a computer display) and other objects from the real environment may appear in the user’s peripheral vision. On the other hand, iVR presents a natural realistic 3D space, typically through head-mounted displays, without interference from real objects in the periphery. Third, the activities performed on a desktop VE, including movement/locomotion (e.g. walking or running), manual or bodily actions (e.g. manipulating objects), are typically accomplished by using a computer mouse rather than directly using one’s body parts (e.g. hands and feet). In iVR, however, the user performs actions and movements by physically engaging his or her own body parts, e.g. through the use of haptic gloves or motion sensors attached to the hands, arms or feet. It is important to note that physical movement or locomotion within the virtual space in iVR may still be restricted by limited physical space, although VR treadmills have been developed to circumvent this problem. Advanced VR technologies such as iVR enable users to manipulate or interact with the environment seamlessly as in real-life situations, providing the greatest amount of immersion and interaction among different virtual platforms. **Fig. 1** (panels **B** and **C**) illustrates an example in which the user picks up a broom in the virtual kitchen while learning the foreign language word for broom.

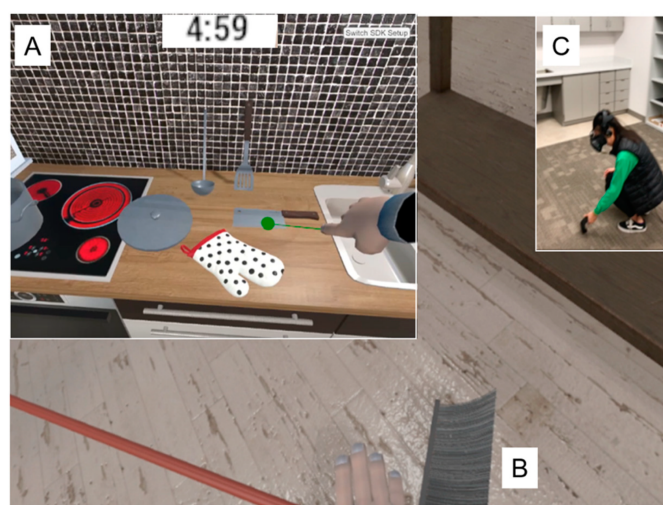


Fig. 1. Interaction in immersive VR (adapted from (13)). (A) Participants could use their handset to point to any item in the virtual kitchen and listen to the corresponding foreign language word; in this example, *dao* (knife in Chinese); (B) individuals could virtually pick up and move any object by pressing a trigger button with their index finger; in this example, a broom; (C) a learner picking up the broom shown in panel B.

Given the abovementioned differences between desktop VE and iVR, it is important to examine the effects of these different platforms on student learning and to explore individual differences in learning. These are discussed in the following sections.

Situated learning and the embodied cognition theory

VR technologies, including both desktop VE and immersive VR, hold great promises for student learning given their immersive and interactive features, as described above. A significant aspect of VE and iVR in the context of student learning is their relation to the learning situation; they provide the contexts – simulations of real-world situations – of learning, that is, “situated learning” (11). By simulating real-life environments so that learning takes place in its natural ecology, these technologies have the potential to enhance educational outcomes (11). Situated learning is rarely consistently applied in traditional classroom teaching practices given the constraints of a typical classroom, for example, limited space restricts the flexibility of simulating different real-world situations across various learning topics. Both VE and iVR rely on highly interactive features which differ significantly from typical classroom teaching that relies largely on verbal interactions between students and the instructor.

Aside from providing naturalistic platforms for student learning, VE and iVR also function as excellent research tools for educators to understand learning outcomes – both successes and failures – as they offer design features that are not available in traditional methods of learning, especially the flexibility for designing learning contexts that vary systematically in crucial environmental characteristics (14, 15). For example, in a real zoo, one cannot easily rearrange animals to different locations to suit a study but such rearrangements can be achieved in iVR through simple programming. Moreover, from a research point of view, the real world contains too many variables that cannot be controlled or accounted for in a study, i.e. confounding variables. In contrast, iVR provides researchers with the possibility of tight experimental control. For example, the environment can be designed to suit the rigour of a study in terms of type of stimuli, presentation mode and timing, i.e. systematic manipulation of key variables. In general, VR provides both “high ecological validity” and “high experimental control,” thereby lending researchers an excellent tool to study naturalistic events in the lab (see (16) for a recent review).

VR learning can be conceptualised in terms of the embodied cognition theory, according to which embodied experiences – experiences grounded in physical/bodily sensation and perception of the properties of the environment – form an integral part of the individual’s (the experienter’s or learner’s) mental representations of concepts, objects and actions (17–19). For true embodiment to occur, the individual needs to integrate various perceptual, sensory and motoric features from the environment, through whole-body physical interaction with the context in which experience or learning takes place. VR provides an excellent platform to enable this type of experience without the learner being physically present in the real-world environment. The embodied cognition theory highlights the “interaction between perception, action, the body and the environment” (18). According to the theory, body-specific (e.g. head, hand or foot) and modality-specific (e.g. auditory, visual or tactile) experiences help build mental representations from the ground up, thus integrating memories of these experiences with the relevant conceptual representations. Once the representations are integrated as such, they can activate the brain’s perceptual and sensorimotor regions when engaged or retrieved in real time (20–

22).

Situated learning, as defined by Dede (11), is important for students’ learning success and depends crucially on embodied experiences. If virtual environments can ground learning in perception and action as real-life situations do, students should become more successful and the learning brain should display corresponding neurocognitive effects and benefits. Below is a summary of evidence for this hypothesis.

Behavioural performance and neurocognitive outcomes

Several studies have examined VR’s efficacy in students’ learning performance (see (11, 12) for reviews). Even in its more primitive form, the use of dynamic 3D figures and images in VR, compared to their static 2D counterparts, can enhance learning outcomes. In several studies, Pani and colleagues (23, 24) demonstrated that the use of and students’ interaction with 3D graphic brain models produced better learning gains in the study of brain anatomy. Interactions with 3D models including rotating views of the brain allow learners to develop better representations and subsequently aid their memory of the spatial relations between anatomical structures. The authors hypothesised that after 3D interactive learning, students may have developed more fine-grained spatial representations in terms of directional relationships between brain structures such as “surround”, “above”, “before”, “behind”. Note that traditional learning of brain anatomy, with 2D figures on a piece of paper or computer screen, limits the students’ understanding because many structures look very different when viewed from different angles/perspectives. For example, as illustrated in Fig. 2, the egg-shaped putamen and the elongated seahorse-shaped hippocampus look very different when viewed from the medial and frontal perspectives. With 3D representations and the ability to rotate and resize these representations, students receive different experiences and therefore different, in this case enhanced, learning outcomes.

Many studies have also shown positive virtual learning effects in other domains. Ketelhut et al. (25) provided evidence that students can develop better problem-solving skills through immersive interfaces compared to paper-and-pencil based curriculum when learning similar contents. Johnson-Glenberg et al. (5) examined high-school students’ chemistry curriculum in both mixed reality (MR) and traditional classroom contexts and found significantly greater learning gains for the MR sessions compared to the small to moderate learning gains in the regular instruction sessions. The authors suggested that increased levels of embodiment and collaboration were the key to MR’s success. In the area of foreign language learning, many studies have adopted the use of desktop-based VE platforms, such as *Second Life* – a popular 3D virtual platform for gaming, to assist learning (see Table 1 in (13) for a review). For example, in Lan et al.’s study (7), American students were trained on a set of Mandarin Chinese vocabulary. Compared to students who learned the same items via computer-based paired associations, those who learned using *Second Life* needed only about half the number of exposures to attain the same level of performance accuracy and showed faster acceleration of learning. A good amount of recent work has focused on promoting virtual language learning with regard to the degree of engagement and student-teacher interaction (26–29). For example, Chen (26) examined the use of *Second Life* in learning English as a second language and found that virtual environment enhanced students’ engagement and promoted collaboration in communication.

Virtual learning not only produces enhanced behavioural perfor-

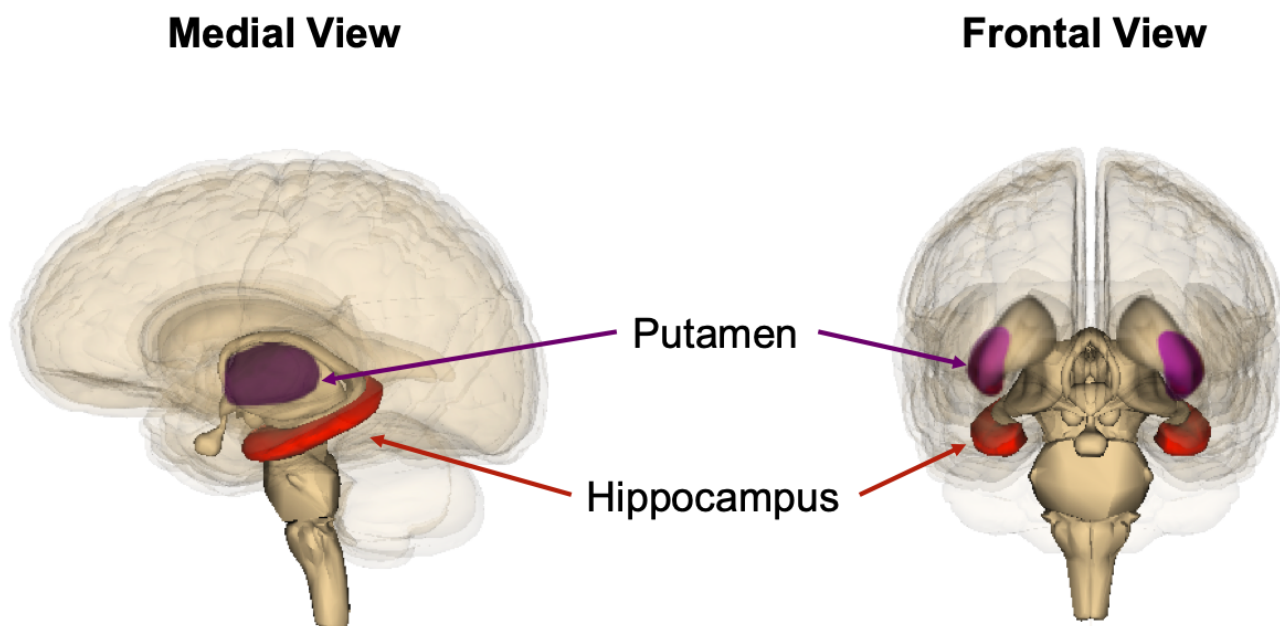


Fig. 2. The putamen (purple) and the hippocampus (red) viewed from medial (left) and frontal (right) perspectives.

mance in many domains (as described above), it also impacts the brain in significantly positive ways. In an earlier seminal study (30), participants navigating a virtual town showed experience-dependent brain changes that were consistent with those associated with real-world navigation experiences. The virtual navigation data led the researchers to identify a network of brain structures that are responsible for human spatial learning; in particular, the right hippocampus was more engaged in allocentric (context- or object-based) representations of space, whereas the inferior parietal lobule (IPL) was more engaged in egocentric (body-centred, self-referential) processes. This dissociation of spatial representations in key brain regions is important for understanding student learning, particularly in the context of the embodied cognition theory. To acquire egocentric representations of space, the learner's immersion in the environment is important and may even be critical. Individual differences in allocentric and egocentric spatial representations will be further discussed below.

Although there has not been a great deal of work investigating the direct association between specific virtual experiences and specific brain changes in language learning, Legault et al. (31), who studied the same participants as Lan et al. (7), observed significantly more neuroplasticity in the right IPL, including both the supramarginal gyrus (SMG) and the angular gyrus (AG), in participants who received VE learning compared to those who learned through paired associations. The authors also noted that cortical thickness – a structural brain measure of gray matter thickness from the surface of the cortex to the white matter – was amenable to rapid change as a result of foreign language learning experience. Specifically, a positive correlation between cortical thickness of the right IPL and training performance was observed across all training sessions in the VE group but the non-VE learners showed a positive correlation in the right inferior frontal gyrus (IFG) – a region associated with effective explicit language training – only at the final training stages (32). Furthermore, cortical thickness in the right SMG was correlated with higher accuracy

score for the VE learners in a delayed retention test, whereas no significant correlation was observed for non-VE learners.

Why would virtual learning of second language vocabulary lead to rapid brain changes in the right IPL? The IPL has been considered the hub of spatial learning (30), vocabulary learning (33), sensorimotor integration and semantic representations (34). In addition, this region has also been associated specifically with effective language learning (35), immersive learning (32) and embodied cognition (36). We speculate that VR learning provides a boost to the IPL's strong engagement in the learner's immersive and interactive activities in virtual environments. In the real world, language learning naturally takes place in spatially enriched environments; for instance, babies begin their native language (L1) learning by listening to their caregivers as they talk and point to objects, while they themselves explore the environment by crawling around. Adults, on the other hand, often learn a second language (L2) in a classroom setting, listening to their teachers talk, reciting lists of L2 words and associating these words with their L1 translation equivalents. Learning in such a context is far removed from real-world environments where language is used in a natural manner and where visuospatial and sensory cues should be present. This lack of real-world experiential support, according to some theorists (e.g. (37)), is the key obstacle to the L2 learner's ability to develop native-like language representations. Recent neurocognitive evidence lends support to this account in terms of the embodied cognition theory. For example, Zhang et al. (38) demonstrated that the neural networks recruited during embodied semantic processing differed significantly between native (L1) and second language (L2) speakers. Specifically, L1 speakers engaged a more integrated brain network connecting key language areas with sensorimotor and semantic integration nodes, whereas the connections between sensorimotor and semantic integration were not strongly engaged in L2 speakers. The less integrated L2 processing system reflects a potential absence of embodied experiences.

VR is uniquely suited to help adults learn languages in simulated environments that maximally resemble real-world contexts, i.e. equating the contexts of child L1 and adult L2 learning. The realistic simulated contexts satisfy situated learning conditions on the one hand and, on the other, provide an opportunity for learners to derive embodied representations that are grounded in social and perceptual interactions (see (39) for a recent discussion of this perspective applied to L2 learning).

Degree of immersion and interaction

Earlier, we identified immersion and interaction as two important features of VR and showed that these two can vary on a spectrum depending on the virtual platforms used (e.g. desktop VE versus iVR). However, it remains unclear whether and how these features independently or jointly contribute to student learning.

Although the examples above illustrate the efficacy of VR in enhancing behavioural performance and brain plasticity, most studies have so far focused on desktop-based VEs. Desktop VEs typically allow little immersion and rely on the use of a personalised avatar – controlled via a computer mouse or remote device – to achieve the necessary interactions. Thus, they generally provide a low to medium degree of immersion and interaction (perhaps with some exceptions from today's gaming industry). In contrast, iVR typically makes use of devices such as head-mounted displays (HMDs), haptic sensors and other body gears (including VR treadmills) to produce “whole-body” simulations in first-person's perspective, allowing users to make social interactions in highly realistic visuo-spatial environments. Given the differences between VE and iVR in terms of immersion and interaction, one would expect the latter which maximises embodied experiences with whole-body engagement to provide better situated learning for students and consequently lead to better learning outcomes.

There have been a number of studies that focus the effects of desktop VE versus iVR on learning. Ruddle et al.'s (40) early study compared the two in terms of navigation patterns with regard to time and efficiency in moving around a virtual environment. The HMD group showed an advantage in both navigation time and the ability to estimate distances. Ruddle and Lessels (41) further showed the importance of full physical movement in VR compared to rich visual scenes for successful memory retrieval performance in navigational search tasks. Another study by Mania et al. (42) tested three groups of participants' recall of objects in a room and found that those who experienced a real room filled with objects showed the best performance, followed by the HMD group; those who viewed the objects on a desktop performed worst. However, after one week, the three groups did not differ in recall accuracy or confidence, suggesting that the differences between the three conditions did not produce a lasting effect. This study is important not only in that it compared iVR with real-world situations but also examined the differential effects on short-term and long-term memory retention. A recent study by Krokos et al. (43) used a within- rather than a between-participants design to examine the efficacy of memory recall in a virtual palace under both desktop VE and HMD iVR conditions. Participants were asked to view and later recall locations of famous faces (e.g. Albert Einstein). Their data showed that HMD iVR, compared to desktop VE, generated significantly higher memory accuracy and fewer errors during recall.

The literature is, however, not uniform with regard to the relative advantages of iVR over desktop VE in producing learning outcomes. While the aforementioned studies have shown better results

with iVR, it is not always the case (44–46). It is therefore important to examine the factors that contribute to the inconsistencies across different studies. Here, we take a different perspective, looking at the individual differences that may have contributed to the inconsistencies (see further discussion below).

The abovementioned studies that have demonstrated the superiority of iVR over desktop VE in terms of aiding learning are generally compatible with the arguments made by Schwienhorst (47) and Johnson-Glenberg et al. (5). Schwienhorst (47) outlined several components of VR that are theoretically important for learning, including increased learner autonomy (e.g. planning and monitoring information) and increased ability of to manipulate and organise the learning environment (e.g. using handheld controllers to move objects within a virtual space). Both autonomy and manipulation are critical for interactions and are in turn important for good performance outcomes in VR. Similarly, Johnson-Glenberg et al. (5) identified three components that are essential for embodied experiences of situated learning: (a) strong engagement of the sensorimotor system (e.g. through movement/locomotion), (b) high degree of interaction with the learning contents (e.g. through gesture or direct manipulation of objects) and (c) learner's perception of being highly immersed (e.g. through a wide-angle 3D view). They showed that mixed-reality (MR) environments involving kinaesthetic and multimodal engagements, compared to traditional instructional methods, led to significant learning gains, given the same teacher, same contents and the same amount of learning time. Although MR (a blend of real and virtual contexts) is not exactly the same as iVR, the degree to which they engage actions and interactions are similar and therefore is believed to enhance learning in similar manners, and thus advantageous over desktop-based contexts where students' bodily engagement is relatively minimal.

Although the comparison between desktop VE and iVR points to the important role of degree of interaction (i.e. the more the better), it remains unclear whether immersion plays an independent role in enhancing student learning performance. Some recent studies suggest that immersion plays a more limited role than one would expect (e.g. (48)). These findings prompt future research directions in which VR technologies should be tailored to individual needs and preferences. The following and final section below centres on the individual difference perspective.

INDIVIDUAL DIFFERENCES IN VR LEARNING

Designers of VR products do not necessarily have access to the cognitive and behavioural characteristics of the users. In addition, average performance does not reveal individual differences among VR users/learners and this may account for some of the discrepancies in the existing literature on the effects of desktop VE and iVR on learning, as discussed earlier. The current paper aims to identify the important features of VR, the characteristics of different learners and the interactions between these two in contributing to learning outcomes – successes and failures. In what follows, we focus on two areas of research to understand individual differences in using VR, namely spatial learning and language learning.

Individual differences in spatial learning

Several studies have demonstrated that learner characteristics – such as cognitive abilities, particularly spatial abilities and executive functions – contribute to individual variability in virtual learning. Here we focus especially on spatial abilities due to the abundant evidence in this domain. Spatial abilities have been examined in

great detail in recent years under the rubric of “spatial learning” (see (49, 50) for reviews; see also (51)), referring to one’s capacity to analyse spatial features of an environment, navigate a complex landscape and form abstract representations of the environment (i.e. mental maps). Various studies have indicated that spatial abilities are central for learning and academic performance in a variety of subjects including science, technology, engineering and mathematics (STEM) (23, 52, 53). A longitudinal study further suggested that spatial ability at age 13 is a strong predictor for later achievement in STEM domains (54).

Research in spatial learning is highly relevant to our discussion because, on the one hand, VR has played a key role in this field (14, 30) and, on the other hand, the issue of individual differences has also gained prominence in this area of research (see (55) for a review). According to Newcombe and colleagues (49, 50), individual differences in spatial abilities can be examined along two dimensions.

First, spatial abilities vary across individuals in terms of how people analyse intrinsic and extrinsic features of the environment. Intrinsic features – the “what” properties of objects – refer to properties such as shape, size and motion patterns, whereas extrinsic features – the “where” properties – concern the relative location and orientation of objects. Interestingly, given the same environment, different individuals tend to excel at attending to different types of features. Accordingly, Motes et al. (56) distinguished between object visualisers and spatial visualisers; these two types of individuals differ in performance and neural efficiency in processing intrinsic appearances of objects and extrinsic spatial relations. Such individual differences have also been linked to achievement in different STEM domains. For example, visual artists and designers – object visualisers – are typically better at intrinsic analytic tasks, whereas mechanical engineers – spatial visualisers – tend to excel at extrinsic tasks (50, 54).

Second, spatial abilities also vary across individuals in terms of how they analyse static representations and dynamic transformations in the environment. Some objects are inherently static (e.g. kitchen appliances such as stovetop and extractor hood), while others are dynamic (e.g. animals that walk, run, swim or fly). To build representations of dynamic objects, the user/learner needs to represent their different movement patterns in the environment (e.g. how animals walk, swim or fly). In VR, researchers can design environments, in a controlled manner, e.g. presenting either static or dynamic objects on separate occasions or both at the same time. Under well-controlled conditions, users can also interact with static objects in specific ways intended by the researchers/designers; for example, they can pick up a broom or open an oven door.

Spatial analyses of intrinsic/extrinsic and static/dynamic features of objects in the environment are natural processes of learning but how these features are differently analysed across individuals may differentially impact student learning. For example, in learning brain anatomy, students need to attend to the shapes of the different brain structures (see above discussion on Fig. 2); for example, the almond-shaped amygdala, the egg-shaped putamen and the seahorse-shaped hippocampus. Students also need to attend to the relative locations of brain structures, e.g. the amygdala is anterior to (or in front of) the hippocampus. As discussed earlier, Naaz et al. (24) examined medical students’ learning of neuroanatomy and found that learning with animated 3D models produced better memory retrieval. Even though the brain’s anatomical structures are static, the use of dynamic 3D models allowed the students to rotate and view the same brain structures from different angles. Using

standardised spatial ability tests (e.g. the mental rotation task) (57), researchers found that students with higher spatial abilities scored significantly higher on neuroanatomy learning outcomes compared to those with lower spatial abilities. However, it remains unclear whether superior performance should be attributed to students’ focus on intrinsic/extrinsic features of the objects or to the treatment of objects in a dynamic fashion, i.e. the opportunity to manipulate and rotate the anatomical structures in a 3D space. Some previous research suggested that participation in full-body movement might be more important than the richness of visual scenes (e.g. (41)) but others have reported contradictory findings. For instance, simply observing how objects were acted on or manipulated, rather than actually performing actions on them, led to better learning outcomes in both real and virtual environments (45).

Studies have demonstrated that human spatial abilities are not uniform across different scales (58); for example, being good at manipulating small objects does not necessarily make one good at navigating a campus (59). Zhao et al. (60) investigated how spatial abilities interact with viewing perspectives in a large-scale environment. Participants navigated a virtual maze through 360° images taken at normal eye level (ground perspective) or 360° images taken at 5.3 metres above ground (elevated perspective). Participants’ spatial abilities were measured using the Santa Barbara Sense of Direction Scale (SBSOD) (61), the scores which have been shown to correlate highly with spatial knowledge acquired from direct experience in large spaces. The results revealed a significant interaction between viewing perspectives and participants’ spatial abilities; in particular, participants with low spatial abilities benefited more from the elevated perspective than participants with high spatial abilities. The results highlight the importance of individual differences in spatial learning and are highly consistent with Legault et al.’s (13) findings in the language learning domain, as discussed below.

Individual differences in language learning

Language learning is closely related to and may depend on spatial ability but this important relationship has often been neglected in the literature. Language learning engages multimodal, perception-action enabled experiences just as spatial learning does (38). Language learning in natural contexts often takes place in rich spatial environments. In the case of babies, they begin learning their first language (L1) while crawling around and exploring their immediate environments. For example, when a mother points to an object and says “teapot”, her child’s attention is instantly drawn to the teapot’s shape, colour and location in the environment. Thus, in language learning, the abilities to mentally analyse and represent spatial features and relations are highly relevant. In contrast, when an adult learns a new word in a second language (L2), he/she makes a mental connection with the equivalent word in L1. For example, upon hearing the teacher say the English (L2) word “teapot” in a classroom, the student’s attempts to mentally connect it to the equivalent word “tetera” in his/her L1, in this case, Spanish. This type of learning is far removed from the natural environment, as the object (teapot) and both its intrinsic and extrinsic features are absent. Here, VR can enhance the L2 learning experience by simulating a natural learning environment through the presentation of visuospatial cues, thereby allowing perception and action to take place.

Previous research demonstrated that physical or bodily simulations such as gestures can promote sensorimotor integration and facilitate both L2 learning (62, 63) and the understanding of

STEM concepts in astronomy, chemistry and earth sciences (5, 64, 65). In cases where there are no word-for-word translation equivalents between L1 and L2, bodily simulations and actions may function as the necessary conditions for understanding the dynamic meanings of certain words. For example, there is no one-word translation equivalent in Chinese for the English word “tiptoe” but simulated movements can easily enhance the understanding of this concept. The close-knit relationship between movement and language processing is supported by neurocognitive evidence. Dynamic movement information processing activates the brain’s sensorimotor regions, damage to which could lead to degraded linguistic performance (66).

While the use of VR in language learning reduces the problems associated with a lack of sensorimotor information, different learners may benefit differentially from the VR experience. Individual differences have been reported in the following studies. Hsiao et al. (67) analysed the movement trajectory patterns of L2 students from Lan et al.’s study (7) which made use of the desktop VE *Second Life* to allow students to explore and learn L2 words in three virtual contexts: a supermarket, a kitchen and a zoo. Students used different strategies in learning new L2 words. In particular the low-achieving learners tended to adopt a more rigid strategy, the “nearest neighbour strategy”, learning labels for objects that were physically adjacent in the virtual space. On the other hand, the high-achieving learners were more likely to adopt a “cluster strategy”, selecting the words that belonged to the same group but were not necessarily nearby. The latter students were more exploratory, grouping similar-sounding words or similar-looking objects together for learning, i.e. grouping objects based on their intrinsic features. Individual differences were captured quantitatively by the statistical method “roaming entropy” – a measure of the degree or variability in movement trajectories in self-directed exploration of space (68). The high-achieving learners showed higher roaming entropy, indicating more variable movement paths and more self-exploratory analyses. Thus, navigation patterns in VE may reflect learners’ capacity in conceptually organising the environment and exploring it interactively.

Although both desktop VE and iVR have gained much attention in language research (see (15) for a review), few published studies have used iVR to examine L2 learning in an experimental setting. Legault et al. (13) were among the first to conduct iVR studies on L2 learning. They trained American monolingual students to learn Mandarin Chinese on an iVR platform. Participants wore head-mounted displays to view and interact with objects/animals in an iVR kitchen/zoo. Researchers compared iVR learning with non-iVR learning method using L2-to-L1 word association (or word-word association; hereafter WW association). Perhaps due to the high degree of both immersion and interaction of iVR features, the iVR group showed significantly better performance in L2 vocabulary attainment than the non-iVR group. Interestingly, the kitchen items were learned better than the animal names. The better performance of kitchen items was attributed to the more action-based manipulations of virtual objects in the iVR kitchen (e.g. picking up objects and moving them around; see Fig. 1) than were possible with virtual animals. The iVR kitchen environment thus conferred more “whole-body” interactive experience to the learner than the iVR zoo, enabling stronger engagement of the sensorimotor system and in turn deriving more embodied representations. Legault et al. (13) also noted a trend that learners with higher spatial abilities performed better in the iVR zoo than in the iVR kitchen. Moreover, the authors reported that the iVR features did not

impact all learners equally. In particular, the “struggling” (or less successful) learners benefitted more from iVR learning. By contrast, iVR learning did not produce a significant advantage over non-iVR learning in successful learners (see Fig. 3). This pattern of individual differences arises from environmental features, learner’s spatial abilities and their interactions. The findings suggest that technologies like iVR can provide a significant boost for some but not necessarily all populations. Finding the sweet spot could be a challenge for future studies.

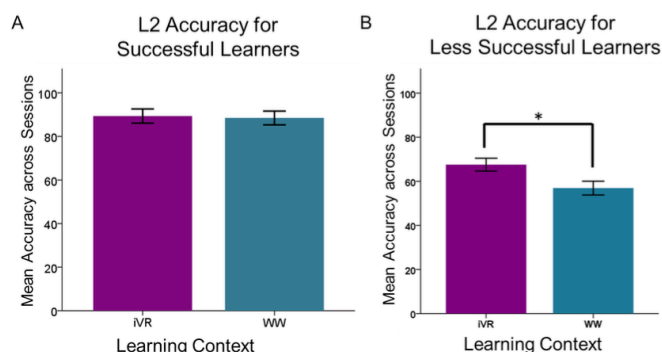


Fig. 3. Effects of Learning Context (iVR versus WW association) on Language Learning (adapted from (13)). (A) There was no significant effect of learning context on successful learners. (B) Less successful L2 learners performed with significantly higher accuracy in the iVR condition compared to the non-iVR WW condition.

Although spatial ability has been the central focus of our discussion on individual differences here, individual differences in executive functions, particularly working memory, could also significantly impact students’ learning under VR conditions. Ample evidence has demonstrated that better working memory facilitates language learning (69–71) but how working memory affects language learning in VR remains unclear. Legault et al. (31) reported that cortical thickness was positively correlated with working memory performance in the right SMG. As discussed earlier, the right SMG – part of the IPL – has been found to play a significant role in L2 learning as part of a larger network in socially grounded language learning (see (39) for review). Interestingly, this correlation was only observed in iVR but not non-iVR learners, suggesting that iVR may bring about differential changes at both behavioural and neural levels across individuals with different working memory capacities.

CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we provided an integrative account of how VR as a rapidly developing technology might significantly impact student learning. Our review has focused on key features of the technology – namely immersion and interaction – as well as important cognitive characteristics of the learners. For VR to play a significant role in education, we need to understand what technological features are important and how students take advantage of the technology. We examined how VR helps to create situated learning conditions and how the embodied cognition theory sheds light in this context of learning. We further presented a summary of how VR learning might lead to enhanced behavioural and brain outcomes compared to traditional methods of classroom-based learning. Our discussion highlighted the importance of individual differences in producing differential learning outcomes under VR learning conditions, in particular how cognitive abilities such as spatial abilities might impact spatial learning and language learning.

Several important issues remain to be understood in future stud-

ies (see (11) for a discussion). First, it is important to understand how the two key features of VR – immersion and interaction – independently impact learning in VR contexts. It may be possible to tease apart their relative contributions by systematically varying the two dimensions in a controlled manner. For example, one could increase the degree of immersion by increasing the FOV and/or 3D image fidelity whilst holding the degree of interaction constant, or vice versa, i.e. increasing the degree of interaction whilst holding the degree of immersion constant. It is also important for future studies to examine the effects of these key features in learners of different age groups and explore the possible interactions. In this regard, further investigation on how individual differences impact VR learning will need to identify other user characteristics that make some learners more receptive to VR learning than others. Second, only a limited number of studies have directly compared VR learning with real-world learning (42, 72). More research making this type of direct comparison is important to determine whether experiences gained in VR will be equivalent or similar to real-life experiences and whether experiences from VR can be transferred to real-life learning situations. For example, if spatial learning and memory performance are enhanced in simulated navigation in VR environments, will the learner fare better in real-world situations? This is a research question that holds significance not only in the context of education and learning but also in other domains including healthcare settings and clinical applications (73, 74). Third, we need to chart in detail the brain networks induced by VR learning across different modality-specific representations. Earlier we pointed out that VR learning could lead to structural brain changes in the inferior parietal lobule but we need more systematic studies of the multiple brain regions and their connectivity patterns induced by VR learning (see (75) for a review of such networks for L2 learning). Regions of interest include, but are not limited to, frontal regions (e.g. superior frontal gyrus – a hub for spatial memory), temporal regions (e.g. anterior temporal gyrus – a hub for semantic representations), pre- and post-central sensorimotor gyri, visual-occipital regions (e.g. MT/V5 – a motion detection centre) and subcortical regions (e.g. basal ganglia – a key structure for sequential learning). Finally, cutting-edge iVR technologies such as VR-eye tracking integration and VR-EEG integration will allow the collection of rich multimodal, real-time, continuous and large-scale data from users' eye gazes/movements, physiological and neurocognitive responses, as well as other bodily actions/movements. Such rich data will provide details about object features that users attend to, spatial movements and navigation patterns that they perform, as well as the duration of their attention and movements on a moment-by-moment basis. These complex data, which differ significantly from data collected from traditional questionnaires and interviews performed after learning, lend themselves readily to data-intensive analytics based on advanced statistical and machine learning techniques. Along this direction, future research may also make use of computational models to unravel learner-specific patterns and further understand individual differences in VR learning.

The abovementioned new and exciting directions for future research will open new avenues for collaborations across different disciplines such as educational science, artificial intelligence and neuroengineering. We stand at a new era of integration of technology and human learning. Technological advancement will allow systematic examination of key features that support VR (e.g. immersive experience, visuospatial perception and motivation), individual differences that affect learning (e.g. cognitive abilities

and executive functions) and the underlying mechanisms (e.g. sensorimotor integration and embodied social interaction) that enable VR as an effective tool for future student learning.

REFERENCES

- Osso VR Chosen as Winner in EdSim Challenge. <https://www.edsimchallenge.com/blog/announcing-edsim-challenge-winner> (2017).
- S. V. Adamovich, G. G. Fluet, E. Tunik, A. S. Merians, Sensorimotor training in virtual reality: A review. *NeuroRehabilitation* **25**, 29-44 (2009).
- K. R. Lohse, C. G. E. Hilderman, K. L. Cheung, S. Tatla, H. F. M. V. der Loos, Virtual reality therapy for adults post-stroke: A systematic review and meta-analysis exploring virtual environments and commercial games in therapy. *PLoS ONE* **9**, e93318 (2014).
- B. M. Brooks, Route learning in a case of amnesia: A preliminary investigation into the efficacy of training in a virtual environment. *Neuropsychol. Rehabil.* **9**, 63-76 (1999).
- M. C. Johnson-Glenberg, D. A. Birchfield, L. Tolentino, T. Kozlupa, Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *J. Educ. Psychol.* **106**, 86-104 (2014).
- H. Jeong, M. Sugiura, Y. Sassa, K. Wakusawa, K. Horie, S. Sato, R. Kawashima, Learning second language vocabulary: Neural dissociation of situation-based learning and text-based learning. *NeuroImage* **50**, 802-809 (2010).
- Y.-J. Lan, S.-Y. Fang, J. Legault, P. Li, Second language acquisition of Mandarin Chinese vocabulary: Context of learning effects. *Educ. Technol. Res. Dev.* **63**, 671-690 (2015).
- B. Dalgarno, M. J. W. Lee, What are the learning affordances of 3-D virtual environments? *Br. J. Educ. Technol.* **41**, 10-32 (2010).
- J. Blascovich, J. Bailenson, *Infinite Reality: Avatars, Eternal Life, New Worlds, and the Dawn of the Virtual Revolution* (William Morrow & Co, New York, 2011).
- M. Slater, Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence* **8**, 560-565 (1999).
- C. Dede, Immersive interfaces for engagement and learning. *Science* **323**, 66-69 (2009).
- M. C. Johnson-Glenberg, Immersive VR and education: Embodied design principles that include gesture and hand controls. *Front. Robot. AI* **5**, 81 (2018).
- J. Legault, J. Zhao, Y.-A. Chi, W. Chen, A. Klippel, P. Li, Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages* **4**, 13 (2019).
- J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, J. N. Bailenson, Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* **13**, 103-124 (2002).
- D. Casasanto, K. Jasmin, *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide.*, A. M. B. de Groot, P. Hagoort, eds. (John Wiley & Sons, Inc, Hoboken, NJ, 2018), pp. 174-189.
- D. Peeters, Virtual reality: A game-changing method for the language sciences. *Psychon. Bull. Rev.* **26**, 894-900 (2019).
- L. W. Barsalou, W. K. Simmons, A. K. Barbey, C. D. Wilson, Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* **7**, 84-91 (2003).
- L. W. Barsalou, *Embodied Grounding: Social, Cognitive, Affective, and Neuroscientific Approaches*, G. R. Semin, E. R. Smith, eds. (Cambridge University Press, 2008), pp. 9-42.
- A. M. Glenberg, M. Sato, L. Cattaneo, L. Riggio, D. Palumbo, G. Buccino, Processing abstract language modulates motor system activity. *Q. J. Exp. Psychol.* **61**, 905-919 (2008).
- L. Aziz-Zadeh, A. Damasio, Embodied semantics for actions: Findings from functional brain imaging. *J. Physiol. Paris* **102**, 35-39 (2008).
- M. Ghio, M. M. S. Vaghi, D. Perani, M. Tettamanti, Decoding the neural representation of fine-grained conceptual categories. *NeuroImage* **132**, 93-103 (2016).
- R. M. Willems, D. Casasanto, Flexibility in embodied language understanding. *Front. Psychol.* **2**, 116 (2011).
- J. R. Pani, J. H. Chariker, F. Naaz, Computer-based learning: Interleaving whole and sectional representation of neuroanatomy. *Anat. Sci. Educ.* **6**, 11-18 (2013).
- F. Naaz, J. H. Chariker, J. R. Pani, Computer-based learning: Graphical integration of whole and sectional neuroanatomy improves long-term retention. *Cogn. Instr.* **32**, 44-64 (2014).
- D. J. Ketelhut, The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in river city, a multi-user virtual environment. *J. Sci. Educ. Technol.* **16**, 99-111 (2007).
- J. C. Chen, The crossroads of English language learners, task-based instruction, and 3D multi-user virtual learning in Second Life. *Comput. Educ.* **102**, 152-171 (2016).
- N. Levak, J.-B. Son, Facilitating second language learners' listening comprehension with Second Life and Skype. *ReCALL* **29**, 200-218 (2017).
- M. Si, A virtual space for children to meet and practice Chinese. *Int. J. Artif. Intell. Educ.* **25**, 271-290 (2015).
- Y. F. Wang, S. Petrino, F. Feng, VILLAGE-virtual immersive language learning and gaming environment: Immersion and presence. *Br. J. Educ. Technol.* **48**, 431-450 (2017).
- E. A. Maguire, N. Bruggess, J. G. Donnett, R. S. Frackowiak, C. D. Frith, J. O'keefe, Knowing where and getting there: A human navigation network. *Science* **280**, 921-924 (1998).
- J. Legault, S.-Y. Fang, Y.-J. Lan, P. Li, Structural brain changes as a function of second language vocabulary training: Effects of learning context. *Brain Cogn.* **134**, 90-102 (2019).
- M. Stein, C. Winkler, A. Kaiser, T. Dierks, Structural brain changes related to bilingualism: Does immersion make a difference? *Front. Psychol.* **5**, 1116 (2014).
- A. Mechelli, J. T. Crinion, U. Noppeney, J. O'Doherty, J. Ashburner, R. S. Frackowiak, C. J. Price, Neurolinguistics: Structural plasticity in the bilingual brain. *Nature* **431**, 757 (2004).
- J. R. Binder, R. H. Desai, The neurobiology of semantic memory. *Trends Cogn. Sci.* **15**, 527-536 (2011).
- J. P. Sheppard, J. P. Wang, P. C. Wong, Large-scale cortical network properties predict future sound-to-word learning success. *J. Cogn. Neurosci.* **24**, 1087-1103 (2012).
- A. McNamara, G. Buccino, M. M. Menz, J. Gläscher, T. Wolbers, A. Baumgärtner, F. Binkofski,

- Neural dynamics of learning sound—Action associations. *PLoS ONE* **3**, e3845 (2008).
37. B. MacWhinney, *Routledge Handbook of Second Language Acquisition*, S. M. Grass, A. Mackey, eds. (Routledge, London, 2012), pp. 211-227.
 38. X. Zhang, J. Yang, R. Wang, P. Li, A neuroimaging study of semantic representation in first and second languages. *Lang. Cogn. Neurosci.* pp. 1-16 (2020).
 39. P. Li, J. Jeong, The social brain of language: Grounding second language learning in social interaction. *Ms. under review* (2020).
 40. R. A. Ruddle, S. J. Payne, D. M. Jones, Navigating large-scale virtual environments: What differences occur between helmet-mounted and desk-top displays? *Presence- Teleop. Virt.* **8**, 157-168 (1999).
 41. R. A. Ruddle, S. Lessels, For efficient navigational search, humans require full physical movement, but not a rich visual scene. *Psychol. Sci.* **17**, 460-465 (2006).
 42. K. Mania, T. Troscianko, R. Hawkes, A. Chalmers, Fidelity metrics for virtual environment simulations based on spatial memory awareness states. *Presence- Teleop. Virt.* **12**, 296-310 (2003).
 43. E. Krokos, C. Plaisant, A. Varshney, Virtual memory palaces: Immersion aids recall. *Virtual Real.* **23**, 1-15 (2019).
 44. H. Li, N. A. Giudice, *Proceedings of the Fifth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, November 5, 2013, Orlando, Florida, USA*, M. Tomko, S. Bell, K. J. Li, eds. (Association for Computing Machinery, 2013), pp. 8-15.
 45. B. S. Santos, P. Dias, A. Pimentel, J.-W. Baggerman, C. Ferreira, S. Silva, J. Madeira, Head-mounted display versus desktop for 3D navigation in virtual reality: A user study. *Multimed. Tools Appl.* **41**, 161-181 (2009).
 46. G. Makransky, T. S. Terkildsen, R. E. Mayer, Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learn. Instr.* **60**, 225-236 (2019).
 47. K. Schwienhorst, Why virtual, why environments? Implementing virtual reality concepts in computer-assisted language learning. *Simul. Gaming* **33**, 196-209 (2002).
 48. J. Zhao, P. Lafemina, J. Carr, P. Sajjadi, J. O. Wallgrün, A. Klippel, Learning in the field: Comparison of desktop, immersive virtual reality, and actual field trips for place-based STEM education. *2020 IEEE Virtual Reality Conference* (2020).
 49. N. S. Newcombe, Thinking spatially in the science classroom. *Curr. Opin. Behav. Sci.* **10**, 1-6 (2016).
 50. D. H. Uttal, D. I. Miller, N. S. Newcombe, Exploring and enhancing spatial thinking. *Curr. Dir. Psychol. Sci.* **22**, 367-373 (2013).
 51. Spatial Intelligence & Learning Center. <https://www.silc.northwestern.edu/research> (2020).
 52. M. B. Casey, R. Nuttall, E. Pezaris, C. P. Benbow, The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Dev. Psychol.* **31**, 697-705 (1995).
 53. M. Kozhevnikov, M. A. Motes, M. Hegarty, Spatial visualization in physics problem solving. *Cogn. Sci.* **31**, 549-579 (2007).
 54. J. Wai, D. Lubinski, C. P. Benbow, Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* **101**, 817-835 (2009).
 55. T. Wolbers, M. Hegarty, What determines our navigational abilities? *Trends Cogn. Sci.* **14**, 138-146 (2010).
 56. M. A. Motes, R. Malach, M. Kozhevnikov, Object-processing neural efficiency differentiates object from spatial visualizers. *NeuroReport* **19**, 1727-1731 (2008).
 57. S. G. Vandenberg, A. R. Kuse, Mental rotations, a group test of three-dimensional spatial visualization. *Percept. Mot. Skills* **47**, 599-604 (1978).
 58. M. Hegarty, D. R. Montello, A. E. Richardson, T. Ishikawa, K. Lovelace, Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence* **34**, 151-176 (2006).
 59. E. Barba, R. Z. Marroquin, Proceedings of the 2017 IEEE International Symposium (Conference Publishing Services, IEEE Computer Society, 2017), pp. 100-110.
 60. J. Zhao, M. Simpson, J. O. Wallgrün, P. Sajjadi, A. Klippel, Exploring the effects of geographic scale on spatial learning. *Cognitive Research: Principles & Implications* **5**, 1-18 (2020).
 61. M. Hegarty, Development of a self-report measure of environmental spatial ability. *Intelligence* **30**, 425-447 (2002).
 62. K. M. Mayer, I. B. Yildiz, M. Macedonia, K. von Kriegstein, Visual and motor cortices differentially support the translation of foreign language words. *Curr. Biol.* **25**, 530-535 (2015).
 63. L. M. Morett, L.-Y. Chang, Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Lang. Cogn. Neurosci.* **30**, 347-353 (2015).
 64. D. M. Markowitz, R. Laha, B. P. Perone, R. D. Pea, J. N. Bailenson, Immersive virtual reality field trips facilitate learning about climate change. *Front. Psychol.* **9**, 2364 (2018).
 65. J. D. Plummer, Spatial thinking as the dimension of progress in an astronomy learning progression. *Stud. Sci. Educ.* **50**, 1-45 (2014).
 66. D. Kemmerer, D. Rudrauf, K. Manzel, D. Tranel, Behavioral patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of actions. *Cortex* **48**, 826-848 (2012).
 67. I. Y. T. Hsiao, Y. J. Lan, C. L. Kao, P. Li, Visualization analytics for second language vocabulary learning in virtual worlds. *J. Educ. Technol. Soc.* **20**, 161-175 (2017).
 68. J. Freund, A. M. Brandmaier, L. Lewejohann, I. Kirste, M. Kritzer, A. Kruger, N. Sachser, U. Lindenberger, G. Kempermann, Emergence of individuality in genetically identical mice. *Science* **340**, 756-759 (2013).
 69. A. Miyake, N. P. Friedman, *Foreign Language Learning: Psycholinguistics Studies on Training and Retention*, A. F. Healy, L. E. Bourne, eds. (Lawrence Erlbaum, Mahwah, NJ, 1998), pp. 339-364.
 70. A. Baddeley, Working memory: Looking back and looking forward. *Nat. Rev. Neurosci.* **4**, 829-839 (2003).
 71. A. Baddeley, Working memory and language: An overview. *J. Commun. Disord.* **36**, 189-208 (2003).
 72. A. Klippel, J. Zhao, K. L. Jackson, P. L. Femina, C. Stubbs, R. Wetzel, J. Blair, J. O. Wallgrün, D. Oprean, Transforming earth science education through immersive experiences: Delivering on a long held promise. *J. Educ. Comput. Res.* **57**, 1745-1771 (2019).
 73. G. Wallet, H. Sauzéon, J. Rodrigues, L. Florian, N. Bernard, Virtual/real transfer of spatial learning: Impact of activity according to the retention delay. *Stud. Health Technol. Inform.* **154**, 145-154 (2010).
 74. G. Wallet, H. Sauzéon, P. A. Pala, F. Larrue, X. Zheng, B. N'Kaoua, Virtual/real transfer of spatial knowledge: Benefit from visual fidelity provided in a virtual environment and impact of active navigation. *Cyberpsychol. Behav. Soc. Netw.* **14**, 417-423 (2011).
 75. P. Li, J. Legault, K. A. Litcofsky, Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex* **58**, 301-324 (2014).

Acknowledgements: **Funding:** The writing of this article was made possible by grants from the National Science Foundation to PL (BCS-1533625) and AK (IIS-1526520) and by a grant from the Guangdong Pearl River Talents Plan Innovative and Entrepreneurial Team (2016ZT06S220). We thank Yingying Peng for her assistance with the preparation and formatting of an earlier version of the manuscript. Jennifer Legault is now at the Department of Linguistics and Cognitive Science, University of Delaware. **Author contributions:** PL conceptualised and wrote the paper; JL, AK, and JZ contributed to literature review, experimental findings and revision of the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper, and/or the materials cited herein. Additional data related to this paper may be requested from the authors.