Virtual Reality (2025) 29:127
https://doi.org/10.1007/s10055-025-01203-3

ORIGINAL ARTICLE

Effects of using depth and motion in virtual reality to enhance Language learning

Mario Vallarino[1] · Rita Cersosimo[2] · Ilaria Torre[1] · Gianni Vercelli[1]

Received: 19 February 2025 / Accepted: 10 July 2025

Abstract

The study explores the integration of a novel virtual reality (VR) approach into language learning within massive open online courses (MOOCs). Specifically, it investigates the effectiveness of complementary VR activities designed to enhance both grammar and vocabulary learning in this context. These activities are intended to be integrated with currently widespread communicative approaches. Starting from Mayer's cognitive theory of multimedia learning, the developed activities embed the depth dimension and body motion into the learning process to enhance its effectiveness. A sample of 40 students was involved in the VR experiment. The learning effectiveness was tested by comparing the outcomes of a group of students participating in the VR experiment with those of another group who carried out comparable screen-based MOOC activities. An additional analysis was made by testing the knowledge of VR group before and after the learning activities. A survey was administered at the conclusion of the post-test to investigate participants' perceptions of the activities in terms of perceived usefulness, perceived learning effectiveness, enjoyment, and comfort. The research results showed promising learning outcomes associated with the novel VR approach in the examined case study. Moreover, the survey analysis provided valuable insights for the design of VR learning activities and revealed a positive reception, which, along with the activities' features, suggests that the approach could be well-suited for integration into language learning MOOCs. The developed framework has been situated within a well-established theoretical model for integrating depth and motion into learning activities in immersive environments, with the objective of enhancing learning performance and experience.

Keywords Virtual reality · Language learning · MOOC · Augmented multimedia principle · Kinesthetic learning

1 Introduction

In recent years there has been growing research interest in the integration of online resources (particularly MOOCs) into face-to-face language learning courses (Conde Gafaro 2019; Jitpaisarnwattana et al. 2019). Such content can be used either as self-consistent resources or as online supplements to in-class courses (blended learning mode). The success of these types of resources is part of the broader picture of the huge success of MOOCs, with the leading platforms, Coursera and Edx, having 118 and 46 million users by 2023, respectively (Shah 2023). At the same time, virtual reality (VR) technologies have been increasingly used in education and in language learning for the past few years. Recently, the world's largest MOOC platform, Coursera, has started to offer courses that integrate VR experiences. In Coursera's Chinese for Beginners course, for instance, there are some experiences that are not mandatory for course completion. In the current landscape of VR usage for language learning,

✉ Mario Vallarino
mario.vallarino@edu.unige.it

Rita Cersosimo
rita.cersosimo@edu.unige.it

Ilaria Torre
ilaria.torre@unige.it

Gianni Vercelli
gianni.vercelli@unige.it

1 Department of Computer Science, Bioengineering, Robotics and Systems Engineering (DIBRIS), University of Genoa, Genoa, Italy

2 Department of Modern Languages and Cultures (DLCM), University of Genoa, Genoa, Italy

Published online: 07 August 2025

there is a variety of approaches, many of which are based on dialogue simulation. However, there is a lack of methods based on a learning theory that incorporates the element that most distinguishes VR from two-dimensional media: the presence of the depth dimension.

The instructional materials developed for this study's experiment leverage VR technology to incorporate depth into the learning process. In addition to sight and sound, which are used in any MOOC language learning course, VR can rely on another powerful element that can intervene in favor of learning: this is motion, which can be used to perceive the depth dimension. The cognitive theory of multimedia learning (Mayer 2020), which will be discussed further on, has studied the most effective ways to use visual information (visual channel) and verbal information (verbal channel) to support learning. With VR, we can use a third channel in addition to the visual and verbal ones: the kinesthetic channel, which is investigated in this work to foster language learning.

The aim of this study is to introduce a novel approach to the use of VR for educational purposes that is both effective and suitable for integration into language learning MOOC courses. The proposed VR approach is designed to enhance grammar and vocabulary as a complement to commonly used methods for developing communicative skills in these courses. Consequently, the first research objective of the present study is to test the effectiveness of this novel approach—which integrates the depth dimension and motion into the learning process—compared to the same activities delivered in a MOOC-based online format. The design of the VR learning activities will be done by applying the augmented multimedia principle, discussed in detail in the Background section, to language learning. The second objective is to investigate the suitability of the approach to the MOOCs model through the statistical analysis of a questionnaire administered to students.

2 Background, theoretical framework, and hypotheses

2.1 VR in education

The use of XR (Extended Reality) technologies now covers a wide range of areas, primarily due to a significant decrease in the purchase prices of such tools (Wu et al. 2019). Extended Reality is used as an "umbrella term" to refer to a set of technologies that, in different ways and degrees, generate virtual elements or environments to be integrated or replaced in the user's physical reality. In the 1990s Milgram et al. (1995) introduced a scheme (depicted in Fig. 1) that, although simplified, still well represents the idea of the continuum between physical reality and virtuality:

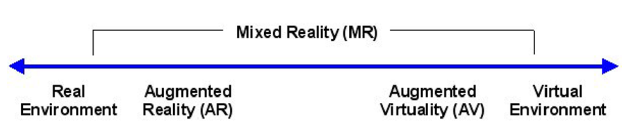

Fig. 1 Reality-virtuality continuum by Milgram et al. (1995)

At one end of the continuum is physical reality, the natural environment in which we are accustomed to living, while at the other end is virtual reality (VR), a completely computer-generated environment. Everything between the two extremes is mixed reality (MR), which involves various degrees of blending between the two ends of the continuum. Thus, augmented reality (AR) introduces virtual elements into physical reality, whereas augmented virtuality (AV) incorporates elements of physical reality into virtual environments. Therefore, extended reality is the overarching category that encompasses the whole spectrum of the continuum, from "completely real" to "completely virtual" (Fast-Berglund et al. 2018).

In education, VR has enormous potential, given its ability to create or reproduce environments in which the learner can move and interact virtually, anticipating real-world situations. VR is widely used in adult training, especially in those areas where it can offer to practice moving safely within dangerous situations, such as military settings or firefighting drills (Xie et al. 2021; Wheeler et al. 2021). In industrial engineering, VR prepares students for real-world challenges by simulating laboratory activities that would otherwise require expensive equipment, which educational institutions are not always able to offer due to lack of funding (Soliman et al. 2021). In the medical field, VR is employed in training to simulate scenarios that future medical professionals will encounter, including surgical procedures (Barteit et al. 2021). Among the many other examples that can be cited in the medical field are certainly the VR environments that allow the exploration of dimensions and positioning of organs and bones, enabling users to dissect and reassemble various parts to better understand anatomy and functional relationships (Jiang et al. 2022). In nursing, one notable application involves simulating a hospital ward complete with patients, medical staff, and family members (Elliman et al. 2016). This kind of simulation allows students to effectively prepare for the responsibilities associated with their role (Plotzky et al. 2021). The use of VR for the improvement of behavioral, communication and social skills of people with disabilities is also gaining ground: studies report how a fun interactive environment facilitates the improvement of skills of this category of people (Bailey et al. 2022; Montoya-Rodríguez et al. 2023). More generally, some studies suggest that VR learning environments tend to be inherently motivating, engaging, and enjoyable for any type of learner, while also promoting content retention

(Al-Azawei et al. 2019; Freina and Ott 2015; Huang and Liaw 2018).

2.2 VR in language learning

In language education, VR has been used in various projects. Frequently, the immersive experience simulates common situations in which a person must use another language to make himself understood (Dhimolea et al. 2022). In a South Florida University study, students were placed in a home-based immersive environment for learning Spanish (Garcia et al. 2019), where they were required to search for an object mentioned in a question. Every time the object was found, a score was received, which contributed to a final ranking (gamification). The results of the study were encouraging, as users reported that they more easily remembered the names of objects that were part of the immersive experience. Furthermore, they also reported that they learned while having fun, thanks to the VR experience and the game dynamics embedded in it. A study by Pinto et al. (2019) compared the delivery of simulations in audio and in VR for learning English: participants were placed in a formal conversation in an office and in another informal one in a pub. The simulations had the same content, which was delivered differently. The simulation results showed a similar level of content retention between the two experiences, but significantly higher participant satisfaction with the VR experience. Alfadil's study (2020) examined the influence of the VR game "House of Languages" on the acquisition of new English words: the research results showed that the experimental group of students immersed in the VR experience was characterized by better learning performance compared to the control group subjected to the traditional method of acquiring new vocabulary. Yudintseva (2024) compared low- and high-immersive VR environments to explore their impact on students' willingness to communicate in English. Participants, working in real human pairs, completed collaborative tasks that required negotiating the arrangement of objects in a virtual space. The instructional methods focused on action-oriented, task-based learning, encouraging authentic communication, collaboration, and decision-making to achieve a shared goal. The findings showed that beyond the level of immersion, these instructional strategies and the sequencing of tasks played a key role in shaping students' communicative engagement. Gruber and Wagner (2024) investigated the use of VR in virtual exchange projects to support intercultural communication and collaboration on environmental issues. Students from the United States and Germany worked in small international groups, using German to address topics connected to the United Nations Sustainable Development Goals (UN-SDGs). The results showed that students developed cultural awareness, with the VR setting supporting authentic interaction in a positive, comfortable environment for meaningful communication.

While several of these studies focus on communicative interaction in immersive settings, the present work proposes a complementary direction by investigating how grammar and vocabulary acquisition can be enhanced through the integration of depth and motion in a VR-based learning process.

2.3 Embodied and kinesthetic approaches to language learning

The efficacy of VR in language learning finds support in theoretical frameworks that underscore the significance of bodily activation during language acquisition. Over the past few decades, cognitive scientists have introduced the concept of embodied cognition (Glenberg et al. 2004; Barsalou 2008), portraying the mind as intricately connected to the body and the environment. Within this framework, bodily sensations and actions play a crucial role in language comprehension (Asher 1977; Glenberg et al. 2011, 2004; Tellier 2008). Evidence suggests that semantic representations are influenced by perceptions and actions (Andrews et al. 2009; Aziz-Zadeh and Damasio 2008; Willems and Casasanto 2011), and sensorimotor input can lead to faster processing after relevant body movements. For instance, as demonstrated by Wilson and Gibbs (2007), the expression "push the argument" is processed more quickly after the simulation of a pushing movement. Neuroimaging studies have further identified connections between brain areas associated with sensorimotor and semantic processing (Boulenger et al. 2009; Desai et al. 2013; Lacey et al. 2012), possibly due to the integration of sensorimotor experience with abstract concepts. These brain areas are particularly activated when we look at manipulable objects, we process action verbs, or we observe the actions of others (Mahon and Caramazza 2008).

When children are in the process of acquiring languages, their inclination to touch objects and engage with the environment becomes evident, proving instrumental in learning new words. However, in educational settings, foreign languages are often taught without incorporating this tactile dimension, relying solely on verbal instruction, memorization, or translation. The differences in instructional methods between native (L1) and foreign (L2) languages underscore the importance of a perception–action enabled context for successful childhood L1 learning compared to the relatively less successful adult L2 learning, which is influenced by the use (or absence) of embodied and immersive environments. Furthermore, conventional classroom-based L2 learning lacks the essential perceptual-visual-sensory features needed to represent vocabulary, such as the shape, size, motion, and location of items to be learned (Parmaxi 2023). In contrast, immersive learning has the potential to engage learners in a natural, embodied, and perception–action-rich context. Recently, some scholars have explored the role of bodily enactment in language learning, suggesting that embodied techniques enhance comprehension and retention compared to form-based approaches (Della Putta and Suñer 2023).

The first who introduced body movements in language learning is Asher (1966), with the Total Physical Response (TPR) method. TPR is a language teaching approach that is based on the premise that learning a second language can be facilitated by associating language with physical actions and responses, mimicking the way children learn their first language. In TPR, learners respond to commands and instructions in the target language through physical actions, creating a dynamic and immersive language learning experience. This approach is particularly effective in early stages of language learning, because it aligns with the natural ways in which individuals acquire their first language through interaction and movement. TPR aims to make language learning engaging, memorable, and enjoyable by tapping into the connection between movement and cognition. The incorporation of bodily responses and gestures helps learners associate vocabulary and structures with actions, enhancing comprehension and retention. It has been effectively used in various language teaching contexts, including those with VR. Subsequently, research on learning languages through body motion has increased, due to the growing popularity of video game consoles and virtual reality headsets. Two noteworthy studies in this domain include Edge et al. (2013), who developed a Kinect game applying TPR principles for language learning without the presence of a human teacher, and Vázquez et al. (2018), who realized a kinesthetic VR system aimed at acquiring Spanish vocabulary by linking words to body movements within an immersive environment experienced through an HTC Vive headset. While Edge et al.'s (2013) results were influenced by significant individual differences, Vázquez et al.'s (2018) findings highlighted a positive effect of kinesthetic experiences in virtual reality on the memorization of new words.

2.4 The cognitive theory of multimedia learning by Richard Mayer

In the broad field of research dedicated to explaining how learning works and how it can be made effective, a special place is held by the studies of the American educational psychologist Richard Mayer. His research focuses on how people learn when they use more than one medium as support (for instance, written text and images in a book, or spoken text and animations in a video).

Mayer's approach to learning is based on a straightforward but not obvious assumption: «A fundamental hypothesis underlying research on multimedia learning is that multimedia instructional messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not» (Mayer 2005).

According to Mayer's research, our mind uses two channels to process a multimedia message, one for verbal information and the other for visual information. These are called the verbal channel and the visual channel, respectively. The verbal channel processes spoken or written words, while the visual channel processes images. In working memory, verbal and visual information are organized into verbal and visual mental models, which are integrated with each other and with prior knowledge to produce learning, as shown in Fig. 2 (Mayer 2010).

The dual-channel model provides the foundation for the well-established *multimedia principle,* anticipated in its core elements since the early 1990s (Mayer and Anderson 1991) and subsequently tested under multiple experimental conditions over the following three decades: "People learn better from words and pictures rather than from words alone" because they can construct verbal and visual mental models, establishing connections between them (Mayer 2020). One of the initial advancements in research on the multimedia principle involved identifying five key principles of multimedia learning (Mayer 2008). According to the *spatial contiguity* principle, learning is enhanced when corresponding printed words and images are presented close together. Based on the *temporal contiguity* principle, individuals learn better when corresponding narration and animation are presented simultaneously. The *coherence* principle states that people learn more when extraneous material is excluded from the presentation. Based on the *redundancy* principle,

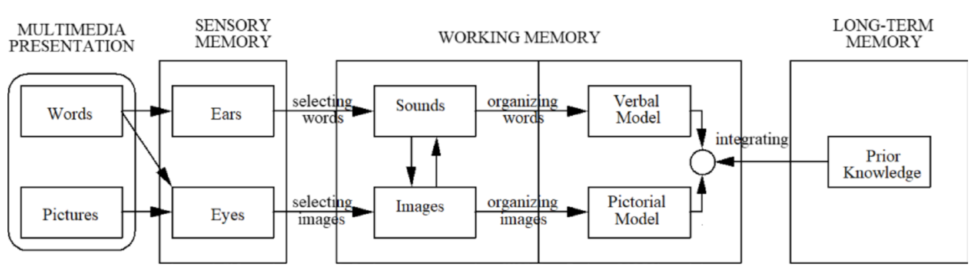

Fig. 2 Cognitive theory of multimedia learning by Mayer (2010)

Fig. 3 Human information processing in two-dimensional media

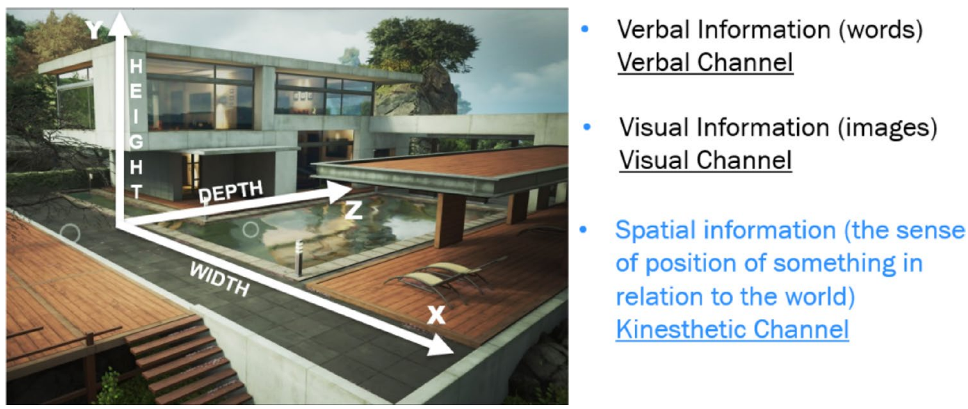

Fig. 4 Human information processing in immersive 3D media

people learn better from animations and narrations rather than from animations, narrations and text (Mayer 2008).

2.5 The multimedia principle in immersive media: exploring the effects of depth perception

According to the cognitive theory of multimedia learning, in two-dimensional media the written or spoken word is processed by the verbal channel, while the image, developed in height and width, is processed by the visual channel (see Fig. 3).

In immersive 3D media, such as VR, the available dimensions become three, and in addition to width and height, we also have depth. Words continue to pass through the verbal channel, and images through the visual channel, but how does the dimension of depth affect our perception?

In a virtual environment, we perceive depth by moving within the scene, utilizing both our muscles and the headset controllers. Exploring the scene, we clearly perceive the position of the elements within the scene itself. Therefore, there is an additional type of information that our brain processes: the spatial information, which concerns the sense of the position of anything within the scene relative to the world we are exploring. Thus, it can be hypothesized that the presence of depth activates a third channel in addition to the verbal and visual channels: this channel can be called the *kinesthetic channel*. The name originates from the word "kinesthesia," which means "sensation of motion" (see Fig. 4).

With the addition of the depth dimension and the consequent activation of the kinesthetic channel, which is supposed to enable the acquisition of spatial information, it is reasonable to assume that the mental models among which connections are established to generate learning become three: verbal, visual and kinesthetic mental model. This may facilitate an even more effective learning compared to that achieved using two-dimensional media.

According to the above, in immersive 3D media, the multimedia principle can be updated as follows: "People learn better from words, images and the position of bodies in three-dimensional space than from words and images alone", because they can construct verbal, visual and kinesthetic mental models, establishing connections between them. We refer to it as *augmented multimedia principle.* It was first introduced in (Vallarino and Vercelli 2023), along with prototype instructional materials developed for French language learning. The augmented multimedia principle is applied when learners are required to physically move their body (kinesthesia) within the three-dimensional virtual space to reach and manipulate virtual objects that have a precise location and are related to a concept to be learned or applied. This feature supports bringing abstract concepts (such as grammar) into three-dimensional space to enhance learning.

2.6 Research hypotheses

Based on the assumption regarding the existence of the kinesthetic channel and the consequent possibility of applying the augmented multimedia principle in immersive learning activities, two types of VR activities have been developed, referred to as “Grammar” and “Vocabulary”. These activities aim to enhance the comprehension of grammatical forms and learn new foreign words, respectively. Through these VR activities, detailed in sections 3.2 and 3.3, we aim to investigate the following five research hypotheses, based on the literature concerning the use of VR for educational purposes.

Hypotheses 1 and 2 Considering that the effectiveness of the multimedia principle has been demonstrated in numerous studies (Mayer and Gallini 1990; Mayer and Anderson 1991, 1992; Mayer et al. 1996; Ponce and Mayer 2014), and that the application of the augmented multimedia principle represents its recontextualization and enhancement in immersive learning environments, it is predicted that VR activities of the Grammar type, implemented by applying the augmented multimedia principle, will be effective in enhancing comprehension of grammatical concepts (Hypothesis 1). Similarly, it is predicted that VR activities of the Vocabulary type, implemented by applying the same principle, will be effective in fostering the memorization of new words (Hypothesis 2).

Hypothesis 3 *Taking into account that these activity types have been designed for widespread integration into online courses offering VR learning experiences, there are some other aspects that are interesting to investigate alongside their learning effectiveness. These aspects are related to the way in which the activities are perceived by students.*

Over time, many studies have investigated students’ presence during VR educational activities (Fowler 2015; Petersen et al. 2022; Makransky et al. 2019; Huang et al. 2020), where ‘presence’ refers to «the psychological sense of ‘being there’ in the environment depicted by the virtual simulation» (Makransky and Mayer 2022). In this regard, Makransky et al. (2019) highlight that VR applications that create a strong sense of presence do not necessarily enhance learning. The authors conclude that considerations regarding the specific opportunities offered by immersive virtual reality for learning should be taken into account in designing educational content for this medium (Makransky et al. 2019). The study by Huang et al. (2020) comes to similar conclusions, highlighting the need to expand the investigation beyond the presence construct. Considering that presence is a controversial element with respect to learning, also revealing little about how participants perceive the VR learning experience, the investigation of the present study focused on four other dimensions, considered more suitable for describing the participants’ experience: perceived learning effectiveness, perceived usefulness, enjoyment, and comfort during the activities, that will be addressed in our 3-5 hypotheses.

Zhang et al. (2017) studied the perceived learning effectiveness of VR and highlighted how it is related to various factors, including appreciation for the learning experience and the number of errors in an assessment test. Perceived usefulness has been defined by Davis (1989) as «the degree to which a person believes that using a particular system would enhance his or her job performance». A study by Ferdinand et al. (2023) examined whether the effectiveness of a science education VR lecture could be enhanced by informing students about the usefulness of VR before the lecture took place. The study demonstrated that a strong perception of the educational usefulness of VR activities enhances the learning achievement. Therefore, Hypothesis 3 in this study is: The perceived usefulness of Grammar and Vocabulary activities will be in a positive relationship with perceived learning effectiveness.

Hypothesis 4 *Enjoyment of learning has been studied by Pekrun et al. (2011) and by Makransky and Mayer (2022) with regard to VR: both studies found strong evidence of the relationship between enjoyment and learning. Consequently, Hypothesis 4 of this study is: The enjoyment experienced in carrying out VR learning activities will be in a positive relationship with perceived learning effectiveness.*

Hypothesis 5 *Many studies investigated discomfort in immersive environments, mainly addressing simulator sickness (Chang et al. 2020; Risi and Palmisano 2019; Bonato et al. 2009; Fernandes and Feiner 2016; Pouke et al. 2018), whereas others (Brunnström et al. 2020; Simone et al. 2019) examine the Quality of Experience (QoE), focusing on comfort aspects deemed relevant for the task at hand. In the present study, we refer to the comfort of VR educational activities as the comfort of an environment where the learner spends a certain amount of time, interacting with some of the furnishings/objects found within that setting. In this regard, the activities carried out are not very different from a simulator, as for example in the study on the comfort perception in a VR simulator of an aircraft cabin, carried out by Ricci et al. (2022): «the passengers’ acceptance of a transportation system, and their willingness to use it again, are strictly related to the aircraft cabin features and spaces,* *which must elicit a sense of comfort and provide a pleasant flight experience». Moreover, «visual stimuli contribute to the formation of personal and peripersonal space perception, thus affecting the spatial perception of comfort, such as legroom, seat pitch, passage width, storage space as well as the sensation of having enough space to work, entertain or rest during the journey. Indeed, recent available research suggests that the visuomotor and visual perception of an environment could influence the experience within the environment itself» (Ricci et al. 2022). The study, carried out using EEG (electroencephalogram) techniques, measured the participants' alpha brain wave activity (frequency band 8–12 Hz) throughout the experiment. The alpha wave is considered a possible marker of stress or arousal condition, since power in the alpha band is higher in the resting state. Depending on the type of stimulus or task, alpha brain rhythm responds with an ERD (Event-Related Desynchronization, alpha power decreases) or an ERS (Event-Related Synchronization, alpha power increases). Specifically, ERD occurs in brain regions involved in task execution, while ERS is observed in regions not relevant to task execution. The results of the experiment demonstrated that the group experiencing a lower level of comfort in the cabin also showed a greater ERD compared to the group experiencing higher comfort. In other words, the lack of comfort proved to be mentally more demanding/stressful than a good level of comfort. The implication of this finding for the present study is that the higher the comfort level in an immersive learning environment, the greater the level of concentration the participant can devote to the learning activity. This is because they can avoid cognitively processing distracting stimuli resulting from suboptimal or poor comfort. Therefore, Hypothesis 5 of this study is: The comfort experienced by learners will be in a positive relationship with perceived learning effectiveness.*

Regarding Hypotheses 4 and 5, while enjoyment and comfort have been examined in prior research on VR learning, we included them in this study to better understand their role in the context of bodily motion and depth-based interaction. Unlike traditional VR designs that often emphasize visual immersion alone, our framework requires learners to engage physically and spatially with the content, introducing potential drawbacks or benefits in terms of ease of use and pleasantness. Therefore, it was important to assess to what extent comfort and enjoyment would support the perceived effectiveness of the learning experience. Hypotheses 1 to 5 are outlined in Table 1.

Table 1 Outline of hypotheses 1 through 5

Hypothesis 1	VR activities of the Grammar type, implemented by applying the augmented multimedia principle, will be effective in enhancing comprehension of grammatical concepts
Hypothesis 2	VR activities of the Vocabulary type, implemented by applying the augmented multimedia principle, will be effective in fostering memorization of new words
Hypothesis 3	The perceived usefulness of Grammar and Vocabulary activities will be in a positive relationship with perceived learning effectiveness
Hypothesis 4	The enjoyment experienced in carrying out VR learning activities will be in a positive relationship with perceived learning effectiveness
Hypothesis 5	The comfort experienced by learners will be in a positive relationship with perceived learning effectiveness

3 Materials and methods

3.1 Participants, device and Procedures

We designed an experiment to test the research hypotheses for the implemented VR activities. The selected case study was an English as a Second Language (L2) course offered by the University of Genoa, available as an online MOOC that incorporates structured activities targeting grammar and vocabulary development. The course was selected because its learning materials were readily available and suitable for experimental use, and because the institution is committed to innovation and the enhancement of its online courses. The grammatical structures and topics covered in the online course align with the requirements of the B1 level, as established by the CEFR (Common European Framework of Reference for Languages), an international standard for assessing language skills. The course is aimed at providing students with a self-study preparation instrument for the English B1 assessment test at the University of Genoa. The VR activities were specifically designed as immersive adaptations of grammar and vocabulary exercises from the English B1 online course. Sections 3.2 and 3.3 will describe them in detail.

The experiment involved 40 students from the University of Genoa, with a background in the humanities but with digital skills proved by having taken exams in ICT subjects. To test learning effectiveness, we compared the outcomes of the 25 participants who carried out the grammar and vocabulary VR activities designed for the experiment ('VR group') with those of the 15 participants who undertook comparable screen-based activities in the University of Genoa's online 'English B1' course ('screen-based group'). The MOOC-based screen condition was selected as the comparison condition against the experimental VR condition due to its instructional continuity and relevance to the intervention, as the VR activities were specifically designed as immersive transformations of the grammar and vocabulary exercises from the same English B1 online course. The two groups differ in sample size due to practical constraints in participant recruitment. Specifically, all 25 students who were available during the scheduled experimental sessions were included in the VR condition. In contrast, although more students initially had access to the online course, only 15 completed all required activities within the time allocated for the screen-based condition. Therefore, group sizes reflect the maximum number of participants that could be involved under the practical constraints of the study setting.

The experimental procedure was designed to address the five hypotheses outlined in section 2.6. As for Hypotheses 1 and 2, the procedure aimed to investigate the effectiveness of the VR activities for grammar comprehension and vocabulary memorization against the MOOC-based activities. In addition, for these hypotheses, the knowledge of the participants in the VR group was assessed both before and after the VR activities through a pre-test and post-test. For Hypotheses 3, 4, and 5, the experimental procedure focused on exploring participants' perceptions of the VR learning experience in terms of perceived usefulness, perceived learning effectiveness, enjoyment, and comfort. An overview of the experimental procedure for the screen-based group and the VR group is provided in Fig. 5. It summarizes the study procedures and shows the sequence of activities for both VR and screen-based groups, as well as the point where learning outcomes were compared.

The screen-based group performed grammar and vocabulary activities on the Moodle platform hosting the course and then took a final test, which was used for comparison with the final test administered to the VR group. In Fig. 5, the latter is referred to as the 'post-test' for the VR group, since a pre-test was also administered to that group—an option that was not feasible within the structure of the university's screen-based course.

Each unit of the online course had a specific theme providing context for the learning activities. The theme of both the Grammar and Vocabulary activities selected for the study was: "We love shopping!". The online course activities follow the programmed discovery technique (Jamieson 1970), which promotes active knowledge construction through guided exploration. After being presented with example sentences that illustrate a specific grammatical structure, students are led through a series of questions designed to help them uncover the underlying rules. These questions include fill-in-the-blank, cloze types, and multiple-choice questions. Subsequently, students are asked to self-assess their knowledge by completing exercises using the same types of questions. All activities followed a sequential structure, where each task became available only upon the successful completion of the preceding one. Participants had four weeks to complete all assigned tasks, with a recommended workload of approximately four hours per week. Further information about the activities in the online course is provided in sections 3.2 and 3.3.

One week after the conclusion of these activities, a final test consisting of 14 questions was administered. The test was administered in a single session in a computer laboratory at the University of Genoa. To ensure a correct comparison of outcomes between the screen-based group and the VR group, it was necessary to convert the final test scores of the screen-based group, which were originally on a 30-point scale, to a 20-point scale. This conversion allowed for an accurate comparison with the VR group's scores, which were assessed on a 20-point scale.

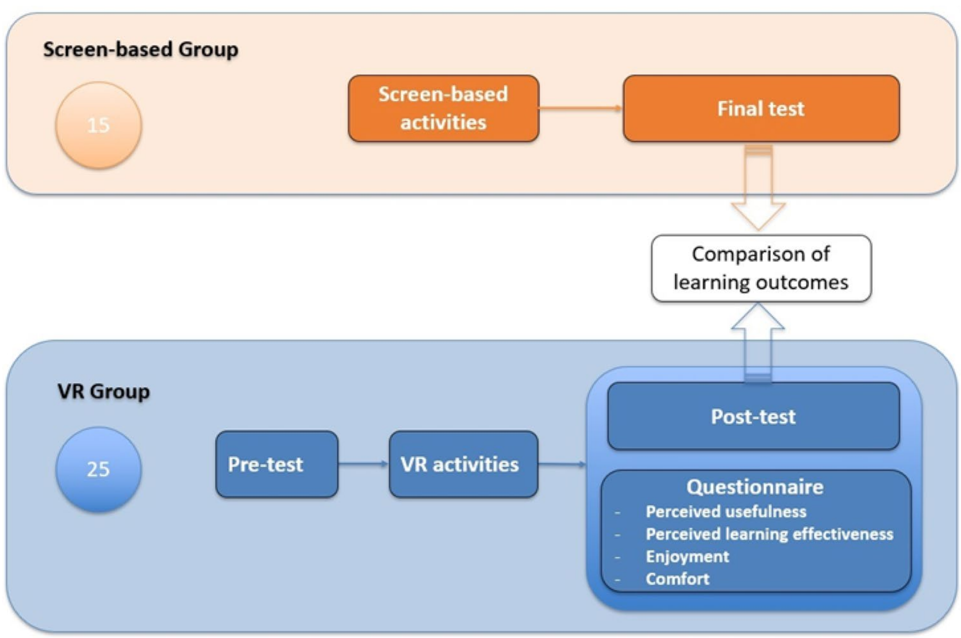

Fig. 5 Overview of the experimental procedure for the screen-based group (top) and the VR group (bottom)

It is worth noting that the primary aim of this study was to compare the activities delivered in the traditional MOOC-based course with their VR adaptation, which was designed according to the augmented multimedia principle. Although the experimental design could have included additional conditions, such as VR language learning environments without depth perception and body motion, this work specifically focused on evaluating the effectiveness of a VR setup that explicitly integrates these features. Such a setup was conceived as a potential innovative tool for integration into formal language learning curricula. Accordingly, the present study compares the learning outcomes achieved through this novel VR approach with those obtained through the existing screen-based online course.

The VR group performed activities specifically developed as immersive, kinesthetic versions of the grammar and vocabulary activities from the 'We love shopping!' unit of the MOOC-based course. These VR activities were designed to support language acquisition while maintaining alignment with the original instructional content. Sections 3.2 and 3.3 provide a detailed description of the design features and instructional principles that informed their development, including the augmented multimedia principle. As explained above, in the VR condition it was possible to deepen the investigation of learning effectiveness by measuring participants' knowledge of the topics covered by the VR activities both before and after the intervention, following the 'pre-test– treatment– post-test' research scheme. Therefore, the final test administered to the VR group (post-test) enabled comparison with the corresponding final test taken by the screen-based group.

Finally, an anonymous survey was administered to the VR group after the post-test to investigate how participants perceived the activities, thereby addressing Hypotheses 3, 4, and 5.

Pre-tests of the VR group took place in a single session during a classroom lecture, whereas some students who were absent arranged for a dedicated appointment. Two sessions of about 60 minutes each were organized for every participant to carry out the 6 implemented VR learning activities individually. The physical space available for motion within the immersive environments of the VR activities was 2.5 x 4 meters. All VR activities implemented for the case study were developed using Unreal Engine 4 software to be experienced through the Meta Quest 2 VR headset.

In the first session, each student underwent an introductory VR activity lasting an average of about 8 minutes, to familiarize with the medium. This was followed by the first set of language learning VR activities designed for the experiment. The Grammar activities ranged in completion time from 5 to 15 minutes, whereas the Vocabulary activities ranged from 3 to 5 minutes. At the end of each of the two sessions students were provided with some short videos on the topics covered in the VR activities. The sessions, as a rule, were held one week apart from each other. The post-test took place one week after the final VR activity session, immediately followed by the administration of the anonymous survey.

Pre-test and post-test consisted of 20 questions to be answered in a maximum time of 20 minutes: 15 questions were about the comprehension of grammatical forms and 5 questions were about the memorization of new words. The questions concerned the learning content of the implemented VR activities. The 15 grammar questions were distributed into 3 questions for each of the five topics covered by the VR activities (Comparisons; Enough and too; Ability, permission, and possibility; Fit and suit; Agreeing and disagreeing); 5 memorization questions consisted of 18 items to memorize, distributed across two identical VR activities in terms of execution, with 9 items each.

The final questionnaire consisted of 15 questions measured on a 7-point Likert scale, aimed at investigating the four dimensions addressed in hypotheses 3-5. To measure these dimensions, we adapted relevant scales from the literature: Davis (1989) for perceived usefulness, Taçgın (2020) for perceived learning effectiveness, Lin et al. (2002) for enjoyment, and Brunnström et al. (2020) for comfort. Appendix A provides the items and the Cronbach's alpha values calculated to ensure the reliability of each scale, as explained in Section 4.3.

3.2 Grammar-type activities

As previously described, each unit of the English B1 online course is built around a specific theme that contextualizes the learning activities. For both the Grammar and Vocabulary activities selected for the VR transformation, the theme was "We love shopping!". This unit includes 17 structured activities designed to develop reading, listening, writing, grammar, vocabulary, and pronunciation skills within the context of shopping. The grammar activities focus on comparatives, quantifiers (*enough*, *too*), modal verbs (*can*, *could*), and expressions of agreement/disagreement (*so do I*, *neither do I*). Vocabulary tasks address clothes, footwear, and accessories. Several activities simulate components of the PET exam, including listening and reading sections, while others target pronunciation practice and sentence correction in writing. The unit concludes with a revision task aimed at preparing students for the PET writing test. All activities are delivered in quiz format and aligned with B1 CEFR descriptors.

A selected subset of the online course activities was identified as suitable for VR transformation. Tasks involving extended reading or writing were excluded, as these activities are more naturally suited to two-dimensional media. In VR, such tasks would require prolonged use of the headset and complex input management through the controllers, which would likely result in physical discomfort and increased cognitive fatigue. These activities are inherently more efficient and ergonomically appropriate when carried out on a traditional screen. Likewise, listening activities were not considered appropriate, as they rely exclusively on hearing and do not benefit from immersive interaction. In contrast, grammar and vocabulary activities were selected, as their transposition in the immersive environment could benefit from the simultaneous engagement of visual, verbal, and kinesthetic channels and lead to more effective learning compared to two-dimensional media. Specifically, grammar and vocabulary activities that do not involve extended reading, writing, or listening appear to be promising candidates for kinesthetic VR transformation.

As examples of VR transformation processes in the online English B1 course activities, the "Comparisons" activity is proposed below for the Grammar type and, in the next section, the "Clothes, footwear, and accessories" activity for the Vocabulary type.

"Comparisons" activity became a VR learning activity consisting of 5 levels, each focusing on a sentence applying grammatical rules concerning comparative forms. In designing the VR implementation, it was decided to set all the educational activities in a shoe store, aligning with the unit's theme "We love shopping!". In order to effectively implement the augmented multimedia principle using the kinesthetic channel, it is necessary to bring the grammatical structures into the three-dimensional space. This can be achieved by making them something "solid" and making their components related to rule comprehension manipulable. Hence, in the VR activity, words became virtual solids in the form of shoeboxes, as shown in Fig. 6.

The learner encounters one or more vacant spaces on the shelf where the sentence is placed. To fill them, they have to select a grammatical form (which can also coincide with a single word) from the counter and "physically" move it into the empty space on the shelf. The learner employs their muscular system to grab the boxes using the controller. When approaching a box that can be grabbed, the controller vibrates. The learner can move through the virtual space of the activity by walking as they would in the real world, or they can choose to use the controllers' teleportation system to shorten the path. Once the box is placed, the learner presses a button on the controller to check the correctness of the answer. In case of an error, all boxes that make up the response options will make a tiny hop, and the learner has to remove the inserted box to replace it with another one. When the answer is correct, the system plays the audio track of the sentence that the learner has just composed. By following this approach, the verbal, visual and kinesthetic channels are all involved in the activity and the brain can produce connections between the different mental models, thereby enhancing the quality of learning. It is worth noting

Fig. 6 Initial state of the "Comparisons" VR activity

that, compared to 2D media, the brain processes additional information concerning the sense of position (spatial information) of the educational objects within the scene. Spatial information is acquired by using the learner's limbs and muscles to move within the scene and manipulate the educational objects. All of this is possible in an immersive medium that leverages the depth dimension for instructional aims.

It is clear that, in such a scenario, the integration between the learner's mental models is not sufficient to enhance comprehension of grammatical concepts, as learning does not only work by trial and error, but the learner also needs to know why a certain answer is correct and the others are not. The use of feedback is important in generating learning (Biber et al. 2011; Winstone and Boud 2022). Results from a study by Fein (2017) highlight that «A text narrative providing the student with additional information about the misunderstood subject matter produces better student performance results, up to 3.4 times better, than a student who did not receive any elaboration feedback (text or multimedia)».

Thus, for grammar activities, two types of feedback were designed to intervene after the learner's response, based on the immediate instructional requirement. After an incorrect response, *addressing feedback* is provided to explain the error and guide the learner toward the correct answer. Conversely, upon a correct response, *success feedback* is offered, providing an explanation of the grammatical rule.

In the described example, the incorrect choice "Expensiver" offers the opportunity to clarify how comparatives are formed for words with three or more syllables. Thus, the addressing feedback becomes: "To form regular comparatives, we always use 'MORE' with adjectives that have three or more syllables (e.g., 'MORE EXPENSIVE')," as shown in Fig. 7. The feedback also partially leverages multimedia principles, appearing as a widget within the immersive learning environment, both as written text and as an audio track spoken by the system.

Upon the correct answer, the system pronounces the sentence. Additionally, the success widget is activated, providing a brief explanation of the grammatical rule used in the sentence (Fig. 8). In this case as well, the feedback is both displayed in the immersive environment and spoken by the system. The widgets feature different icons depending on their purpose: an "i" for addressing feedback and a checkmark for success feedback.

The grammar exercises take place in different areas of the store, with differences in box color and shelf height, aimed at diversifying the visual and kinesthetic stimuli used to apply the augmented multimedia principle. An example is shown in Fig. 9.

3.3 Vocabulary-type activities

The activity of learning new vocabulary also relies on the augmented multimedia principle. In this case, the

Fig. 7 Addressing feedback

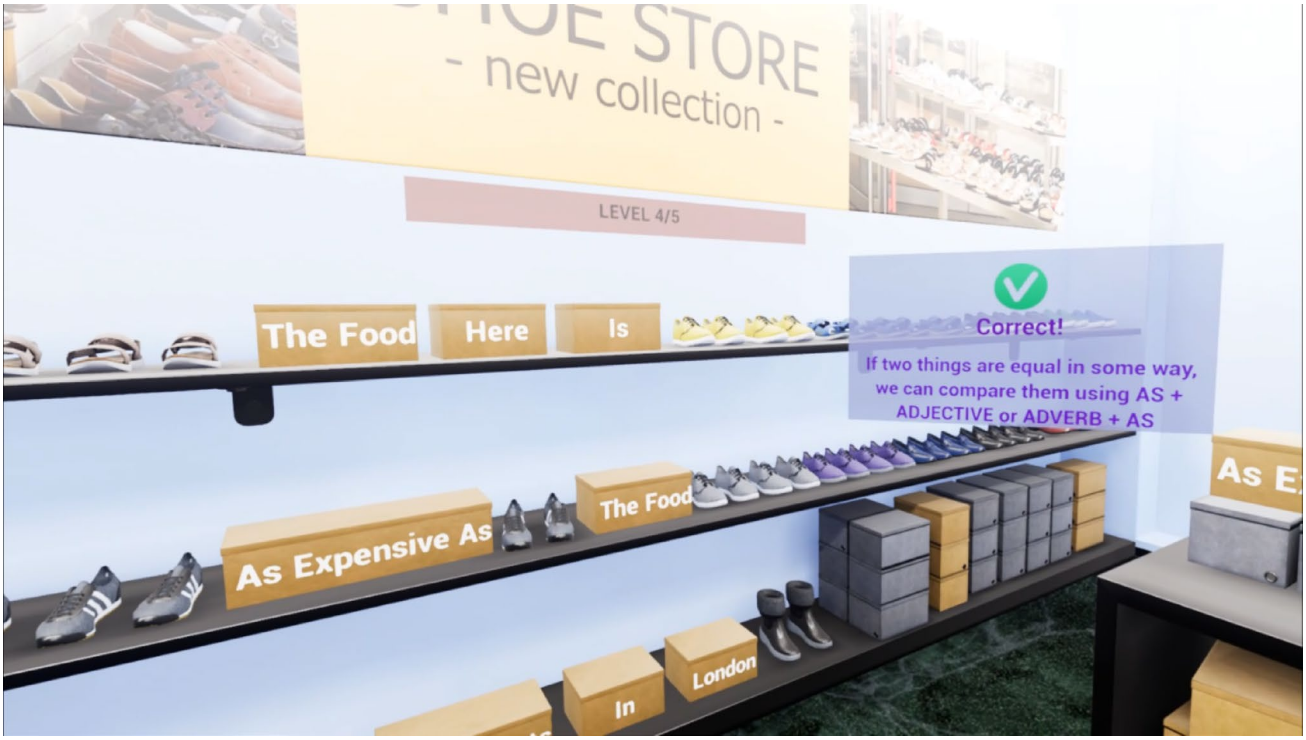

Fig. 8 Success feedback

Fig. 9 Initial state of the "Agreeing and disagreeing" VR activity

transformation into VR was applied to the "Clothes, footwear, and accessories" activity of the English B1 online course. In the current online course format, the activity involves displaying images of 18 objects and selecting their precise names from a dropdown menu. The objects are grouped into sets of three. Given the current state of development in online archives of three-dimensional objects for VR, sourcing 18 distinct objects with the necessary consistency (for instance, regarding mesh definition) would have been a challenging task. Instead, a viable compromise was to include widgets representing the objects to be memorized within the learning activity scene (Fig. 10).

The scene features three widgets, each representing an object, positioned at different heights and angles. Three

Fig. 10 Initial state of the "Vocabulary—clothes, footwear and accessories" VR activity

Fig. 11 Attaching plates to widgets

plates, resembling license plates, are placed on a table, each labeled with the name of one of the three objects. The learner selects one plate at a time and moves within the room to locate the corresponding widget to attach the plate to (Fig. 11).

If the plate is incorrect, it will not attach to the widget when brought close. Conversely, if the plate is correct, it will attach to the widget, and the learner will hear from the system how that word is pronounced. In this type of activity as well, the learner uses their muscular system to "touch" the virtual educational objects (via the controller's vibration), grab them, and then move them to the correct place, engaging all three channels (visual, verbal, and kinesthetic) used by our brain to process information.

4 Results

4.1 Mann–Whitney U test comparing the screen-based group's final test and the VR group's post-test

In Hypotheses 1 and 2, we predicted that VR activities of the Grammar and Vocabulary types would be effective in enhancing comprehension of grammatical concepts and fostering memorization of new words, respectively. As a whole set, the VR activities were compared with analogous two-dimensional activities carried out in the online English B1course by the screen-based group. Given that the scores of the VR group's post-test were not normally distributed according to the Kolmogorov-Smirnov test, a

Mann-Whitney U test was conducted to compare the scores between screen-based group (n = 15) and VR group (n = 25). The test results indicate that there is a statistically significant difference in scores between the screen-based group (M = 13.72, Mdn = 13.60, SD = 4.52) and the VR group (M = 17.68, Mdn = 17, SD = 1.65), U = 89, Z = -2.76, p = 0.006. Specifically, the scores in VR group tend to be higher than those in screen-based group. The result of this comparison supports the effectiveness of VR learning activities.

4.2 Wilcoxon signed-ranks test for pre-test and post-test scores on enhancing grammatical concept comprehension (Grammar-type activities)

Kolmogorov–Smirnov test of observed values and visual inspections of the histograms of the pre-test and post-test scores for the Grammar and Vocabulary activities showed that, in all the cases, scores were not normally distributed. Therefore, we used the Wilcoxon signed ranks test to analyze data of both Grammar and Vocabulary activities. The descriptive statistics for the Wilcoxon signed-ranks tests of the pre-test and post-test scores for the Grammar and Vocabulary activities are reported in Table 2.

We predicted that Grammar-type VR activities, implemented by applying the augmented multimedia principle, would be effective in enhancing comprehension of grammatical concepts (Hypothesis 1). The Wilcoxon signed-rank test was used to test Hypothesis 1, by comparing the pre-test and post-test scores. The test results indicate that there is a statistically significant difference in scores between the pre-test (Mdn = 12, n = 25) and the post-test (Mdn = 14, n = 25), Z = -3.853, p < 0.001. Specifically, post-test scores after doing the VR activities aimed at enhancing grammatical concept comprehension tend to be higher than pre-test scores. With a large effect size (r = -0.82), it can be concluded that test scores significantly vary in the population after completing the grammar VR activities. This result supports Hypothesis 1.

4.3 Wilcoxon signed-ranks test for pre-test and post-test scores on new words memorization (Vocabulary-type activities)

Similarly to the above case, we predicted that the Vocabulary-type VR activities, implemented by applying the augmented multimedia principle, would be effective in fostering memorization of new words (Hypothesis 2). The Wilcoxon signed-rank test was used to test Hypothesis 2, by comparing the pre-test and post-test scores. Also in this case, the Wilcoxon signed ranks test results indicate a statistically significant difference in memorization test scores between the pre-test (Mdn = 2, n = 25) and the post-test (Mdn = 4, n = 25), Z = -4.281, p < 0.001. In particular, the post-test scores after completing the VR activities for memorizing new language words tend to be higher than the pre-test scores. In light of the large effect size (r = -0.86), it can be concluded that the post-test scores show a significant improvement in the population after completing the memorization VR activities. This result supports Hypothesis 2.

4.4 Analysis of student questionnaires

At the conclusion of post-test, each student was asked to complete a brief anonymous questionnaire consisting of 15 questions. All responses were measured on a 7-point Likert scale. The questions aimed to investigate four dimensions: perceived usefulness, perceived learning effectiveness, enjoyment, and comfort of the VR activities. After analyzing the questionnaire scales using Cronbach's alpha, two questions were discarded as they significantly lowered the alpha. Consequently, perceived usefulness was measured by three items adapted from Davis (1989), perceived learning effectiveness was measured by three items adapted from Taçgın (2020), enjoyment was measured by three items adapted from Lin et al. (2002), and comfort was measured by four items adapted from Brunnström et al. (2020). The full list of items, along with the Cronbach's alpha values for each scale, is reported in Appendix A, whereas the descriptive statistics of the four questionnaire scales are shown in Table 3.

Based on the Kolmogorov-Smirnov test and visual inspection of the histograms of the four scales, none of the samples was normally distributed. Therefore, we used

Table 2 Descriptive statistics of pre-test and post-test for the Grammar and Vocabulary activities

	N	Mean	Standard deviation	Minimum	Maximum	Median
Pre-test						
Grammar	25	11.96	1.95	7	15	12
Vocabulary	25	1.72	1.10	0	3	2
Post-test						
Grammar	25	13.92	1.04	12	15	14
Vocabulary	25	3.76	1.09	2	5	4

Table 3 Descriptive statistics of the questionnaire scales

	N	Mean	Standard deviation	Minimum	Maximum
[UTI] Perceived usefulness	25	6.15	.93	3.33	7.00
[EFF] Perceived learning effectiveness	25	5.93	1.25	2.67	7.00
[ENJ] Enjoyment	25	6.41	.82	4.33	7.00
[CON] Comfort	25	6.00	.95	3.25	7.00

Table 4 Correlation coefficients between the questionnaire scales

Construct	UTI	EFF	ENJ	CON
UTI	1.000	.817**	.692*	.311
EFF	.817**	1.000	.852**	.404*
ENJ	.692**	.852**	1.000	.387
CON	.311	.404*	.387	1.000

*p<.05, **p<.01

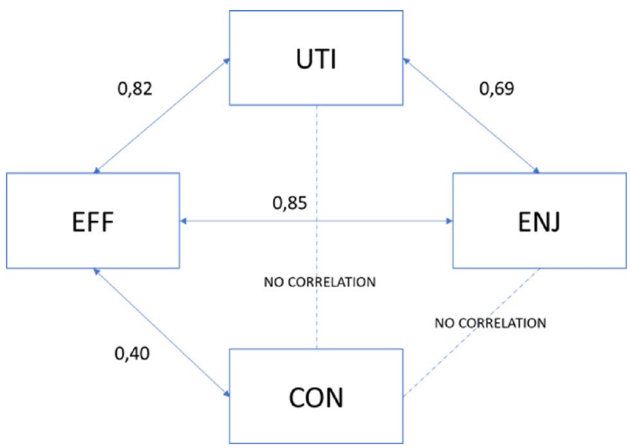

Fig. 12 Diagram of statistically significant correlations

the Spearman's Rho test to detect correlations between the questionnaire scales (the test results are reported in Table 4). Statistically significant correlations between the four questionnaire scales are depicted in Fig. 12, whereas the scatter plots of single correlations are shown in Fig. 13.

.

We predicted that perceived usefulness (Hypothesis 3), enjoyment (Hypothesis 4), and comfort (Hypothesis 5), respectively, would be in a positive relationship with perceived learning effectiveness. The Spearman's Rho correlation matrix highlights a very high positive correlation between perceived usefulness and perceived learning effectiveness (r = 0.82, p < 0.001), which supports Hypothesis 3. A very high positive correlation also emerges between enjoyment and perceived learning effectiveness (r = 0.85, p < 0.001), which supports Hypothesis 4. Finally, a medium positive correlation arises between comfort and perceived learning effectiveness (r = 0.40, p = 0.045), which supports Hypothesis 5. Beyond the scope of research hypotheses, a high positive correlation is also found between enjoyment and perceived usefulness (r = 0.69, p < 0.001), while no correlations were observed between comfort and perceived usefulness, nor between comfort and enjoyment.

5 Discussion

In this section we discuss the results from the experiment in relation to our research hypotheses. The next section addresses the study's limitations and proposes directions for future research. Finally, in the Conclusion section, we

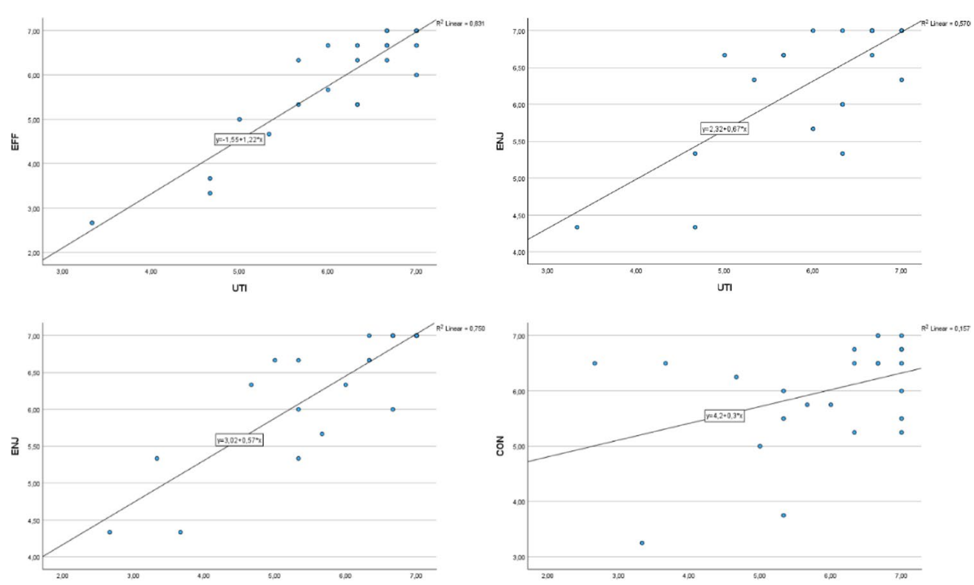

Fig. 13 Scatter plots of individual correlations between the questionnaire scales

provide considerations regarding future prospects, particularly for MOOC courses incorporating VR activities, based on the findings of this study and related literature.

Regarding VR activities, the VR group outperformed the screen-based group in the comparison of learning outcomes, and research hypotheses 1 and 2 were confirmed. In addition to achieving higher post-test scores, participants in the VR group also showed a lower score dispersion (SD = 1.65) compared to the screen-based group (SD = 4.52). Such reduced variability in the VR group may be due to the supportive design of the immersive activities, which helped learners stay focused on key content, provided immediate feedback, and facilitated interaction with the material in a meaningful and engaging way. Therefore, the main finding of this study is that the VR grammar and vocabulary learning activities, designed according to the augmented multimedia principle, were associated to better learning outcomes. This was observed within a short-term timeframe, corresponding to the weekly schedule of a MOOC course. The finding supports the multimedia principle (Mayer 2020) and its updated three-dimensional version, as addressed in (Vallarino and Vercelli 2023) and experimented in this work. Experimental results also align with Petersen et al.'s study (2022), which examines the application of the Cognitive Affective Model of Immersive Learning (CAMIL) to VR learning. According to the CAMIL, two technological VR features, interactivity and immersion, influence a range of cognitive and affective variables that can either facilitate or hinder learning. One such variable is embodied learning, which can be defined as the «pedagogical approaches that focus on the non-mental factors involved in learning, and that signal the importance of the body and feelings» (Panigagua and Istance 2018). As discussed in Section 2.3, embodied learning is related to the approach discussed in this study, but it is worth noting that, so far, various authors have addressed embodied or kinesthetic learning without systematically integrating it into a comprehensive learning theory dealing with the relationships between human information processing, the use of media to support learning, and body motion. Returning to the study by Petersen et al. (2022), the results unexpectedly revealed a negative correlation between learning and the levels of embodied learning expressed by participants. The authors' explanation came from another study by Lindgren and Johnson-Glenberg (2013), who investigated embodied learning and mixed reality. They argued that, for effective learning to occur, participants' actions must be related to the learning content. Consequently, Petersen et al. (2022) pointed out that in their study «it was not possible to perform actions explicitly related to the learning content, which is why the experience of actively using one's body for learning might have done more harm than good». Therefore, they concluded that «it can be theorized that this negative relation stemmed from the lack of congruency between bodily actions and learning content. Hence, an important direction for future research in the field could be to *examine the process of learning using simulations actively designed for embodied learning*» (Petersen et al. 2022). The italicized portion fairly accurately describes what was done in the present study: implementing VR learning activities specifically based on kinesthetic action performed on learning content in the form of virtual objects, proved effective for learning purposes, thus supporting the conclusions drawn by Petersen et al. (2022) and Lindgren and Johnson-Glenberg (2013). Hence, although more research is needed to determine the effectiveness of the augmented multimedia principle in other learning contexts and in long-term retention, the learning outcomes of the present work appear encouraging.

The VR activities experimented in this study align with Mayer's cognitive theory and its kinesthetic extension. To contextualize the value and possible exploitation in learning activities, it is helpful to discuss the distinctive features of this approach alongside those of established language teaching methods, such as Communicative Language Teaching (CLT) and Task-Based Language Teaching (TBLT). CLT emphasizes the development of communicative competence through interactive activities that prioritize meaning over form, such as dialogues, role-plays, and group discussions (Savignon 1972). TBLT, while rooted in similar principles to CLT, organizes learning around real-world tasks that require learners to use the language as a tool to achieve specific outcomes—such as planning a trip, solving a problem collaboratively, or writing a formal email—thereby fostering authentic communication in goal-oriented contexts (Ellis 2021). The distinctive feature of the immersive approach developed in this study is that it engages spatial cognition and sensorimotor processing, potentially fostering deeper multimodal encoding and context-dependent memory recall through embodied interaction. Therefore, it is worth noting that the VR learning framework proposed in this work is not intended as a replacement for CLT and TBLT, but rather aims to investigate the potential contribution of the kinesthetic dimension to the learning process within the context of the developed VR activities (Makowski et al. 2017). This dimension is not necessarily a high-tech approach and can also be incorporated into traditional classroom activities, as discussed in section 2.3. However, VR environments enable a broader and more versatile range of kinesthetic learning experiences.

Based on the results of the post-test survey, most participants enjoyed the VR learning activities, finding them useful, comfortable, and educationally effective. According to Hypotheses 3, 4, and 5, the positive correlations between perceived usefulness, enjoyment, and comfort, respectively, with perceived learning effectiveness were confirmed.

Regarding perceived usefulness, this study supports the findings of Ferdinand et al. (2023), that found an enhancement in the learning achievement when the usefulness of VR activities is perceived. Concerning enjoyment, the research results align with studies by Pekrun et al. (2011) and Makransky and Mayer (2022), which highlight the strong relationship between the enjoyment of performing learning activities and learning. With respect to comfort, the study by Ricci et al. (2022) examined the effects on the brain of perceived comfort in an airplane cabin simulator and found that a lack of comfort in a simulated environment is mentally challenging and activates areas of the brain normally used for task execution. The positive relationship observed in this study between comfort and perceived learning effectiveness supports the findings of Ricci et al. (2022), as a good level of comfort in VR activities reduces the distracting stimuli coming from the virtual environment design, thus enabling the learner to direct greater focus on the learning task. The positive relationship between comfort and perceived learning effectiveness also aligns with Mayer's cognitive theory of multimedia learning, which states that the challenge of instructional design is to guide the appropriate cognitive processing of the learner during learning without overloading their working memory capacity (Mayer 2020, p. 50-51). Extraneous processing involves cognitive processing that is not useful for reaching the learning objective and is caused by poor instructional design (Mayer 2014). The results of the present study regarding comfort, supported by the aforementioned research (Ricci et al. 2022; Mayer 2014), emphasize the importance of comfort in VR learning activities: poor comfort results in extraneous processing for the learner, whereas adequate comfort in activities, limiting extraneous processing from non-functional elements of the virtual environment, can correspond to better perceived learning effectiveness.

The statistical analysis of the survey also revealed an unexpected but strong positive relationship between enjoyment and perceived usefulness. The same positive relationship was also observed in several other studies on VR, including (Lee et al. 2019; Jo and Park 2023; Sagnier et al. 2020). Many studies suggest that since virtual environments are perceived as hedonistic by users, the enjoyment experienced while using them significantly contributes to the usefulness that users attribute to this medium. However, in the context of VR for learning purposes, this explanation alone does not appear sufficient to justify such a strong positive relationship. The study by Jo and Park (2023) provides a more in-depth explanation of the relationship between enjoyment and perceived usefulness of VR. The authors developed a model to identify the elements that most influence the user intention to continue using VR, which includes the dimensions of perceived usefulness and perceived enjoyment. The research results showed that one particular element, the physical affordance, positively influences both perceived usefulness and enjoyment. As explained by the authors, physical affordances refer to «those design features of an interface or system that amplify or simplify physical interactions» (Jo and Park 2023), akin to how the "door handle" affordance enables us to effectively interact with a door to open or close it in the physical world. According to the authors, a system that effectively integrates physical affordances makes it easier to navigate, understand what can be done, and extract value from the system. This increases «the perception of the tool's utility, as users can seamlessly integrate it into their tasks without unnecessary cognitive disruptions» (Jo and Park 2023). At the same time, a system offering remarkable physical affordances makes the user experience more engaging, enjoyable, and satisfying. As the authors emphasize, «when users engage with interfaces that offer favorable physical affordances, they are likely to derive both utility and pleasure» (Jo and Park 2023). Beyond the hedonistic component, which is likely to be present in a well-designed VR learning activity, in the case of the present study it is possible that the significant physical affordances offered by the VR learning activities, consisting of the abilities of "touching" (through vibration), holding "in hand" and moving the objects involved in the learning process, influenced both perceived usefulness and enjoyment.

Hence, the research findings highlight the importance of integrating valid physical affordances and ensuring good comfort in VR learning activities to enhance perceived usefulness, enjoyment, and perceived effectiveness. This can serve as a hint for designing VR activities for language learning MOOCs and promote greater acceptance of the novel approach to VR language learning addressed in this research work.

Lastly, although this study did not include a formal cost–benefit analysis, it is worth noting that the VR activities were developed using freely available software (Unreal Engine 4) and tested on a consumer-grade standalone headset (Meta Quest 2), which has become increasingly affordable for educational institutions. The modular nature of the activities and their short duration also support flexible integration into existing MOOC structures, potentially maximizing educational impact while minimizing implementation cost. Future research could investigate in greater detail the return on investment of this approach, considering both learning gains and the costs of development, deployment, and maintenance.

6 Limitations and future directions

This study presents certain limitations concerning sample characteristics, the short-term focus of the learning assessment, the absence of a qualitative component, and the design of the comparison condition. The sample size was relatively small and demographically homogeneous, as participants were all university students from a specific disciplinary background. Although the results were statistically significant and supported by large effect sizes ($r = -0.82$ and $r = -0.86$), the lack of a broader and more heterogeneous sample limits the generalizability of the findings. The number of participants was determined by the maximum availability of eligible students within the course offering at the time of the study. For this reason, no a priori power analysis was conducted; however, the reported effect sizes provide a robust indication of the magnitude of the observed learning gains. Future studies involving larger and more diverse populations will enable power-based sample planning and support broader external validity.

Another limitation concerns the learning assessment, which focused exclusively on short-term outcomes. While this choice was consistent with the weekly pacing of the course modules, it does not allow for evaluating the persistence of learning over time. Including delayed post-tests in future research would offer valuable insights into the potential of immersive kinesthetic environments to support long-term retention.

A further issue is the exclusive reliance on quantitative instruments to evaluate both learning outcomes and user experience. Although the Likert-scale questionnaire provided structured data on perceived usefulness, enjoyment, comfort, and learning effectiveness, adding a qualitative analysis could reveal more nuanced or unexpected aspects of learners' experience. Future research should consider integrating open-ended questions or semi-structured interviews to gain qualitative insights, particularly to explore how learners engage with the kinesthetic features of VR activities.

Finally, the design of the comparison condition introduces an additional interpretive constraint. Since the screen-based group used traditional 2D materials, while the VR group engaged with a virtual environment involving depth and motion, it is not possible to determine to what extent the observed learning gains are specifically attributable to the depth and motion features, as opposed to the immersive medium itself. Although the choice to compare the full-featured VR condition to a screen-based learning environment was grounded in both the instructional goals and the institutional context, it does not allow for a clear separation between the effects of the overall VR experience and those of its kinesthetic and spatial components. Future studies could address this limitation by including a third group using VR without depth and motion features, in order to disentangle the relative contributions of each factor.

Addressing these limitations in future research will help strengthen the empirical grounding of the study's framework and enhance its applicability across diverse educational settings.

7 Conclusion

As highlighted in the literature review in the Background section, systematic instructional design approaches using VR in the context of language learning have yet to emerge. In MOOC courses, VR is being introduced with optional experiences in addition to the structured course content. The main difference from a traditional MOOC is that a VR-enhanced MOOC includes immersive learning activities conducted in the three dimensions of physical space, rather than in the typical two dimensions of a computer or smartphone screen. The noticeable leap lies in the medium's features, transitioning from 2D to 3D, and involving a Head Mounted Display (HMD) device, which in this study was the Oculus Quest 2.

In two-dimensional media, Richard Mayer's cognitive theory of multimedia learning, which is centered around the multimedia principle, has long proven its effectiveness. The idea developed in this work consisted in updating the multimedia principle to three-dimensional media, by exploiting the dimension of depth to enhance the learning process. The resulting augmented multimedia principle guided the design of two types of VR learning activities, tailored to the case study of an English B1 online course. The short duration and content granularity of these activities make them suitable for applying mastery learning (Bloom 1968). These aspects make it easy to integrate them into language learning MOOC courses (Tang 2017).

The overall results of the experiment suggest that the two types of activities developed, based on the application of the augmented multimedia principle, are suitable for inclusion in language learning MOOCs that offer VR content. Moreover, the survey revealed some guidelines to consider when designing these new types of immersive learning activities. Integration within MOOC courses appears to be straightforward due to the granularity of the learning objectives and their short duration. The implemented activities are complementary to the currently popular VR dialogue simulations, meaning they can all coexist within an online course and provide diversified benefits for student learning.

From a theoretical standpoint, the experimental results provide preliminary evidence supporting the application of the augmented multimedia principle in the case study examined in this work. However, further research is necessary to isolate the specific contribution of depth and motion features from the general effects of the VR environment, and to confirm its validity in language and other learning contexts. Nevertheless, the study has laid the theoretical foundations and analyzed practical evidence for an approach to VR in learning processes that leverages its main advantage over traditional media: the perception of depth through body motion. The described approach led to an update of Mayer's multimedia principle, which emphasizes the importance of the kinesthetic aspect in the perception of the position of objects that are part of an educational activity in an immersive three-dimensional environment, with the aim of improving the effectiveness of the learning process.

Appendix A

Construct	Item code	Item	Source
[EFF] Perceived usefulness	UTI_1	Carrying out the VR activities was a useful review for me	Davis (1989) Cronbach alpha = 0.757
	UTI_2	Carrying out the VR activities helped improve my English	
	UTI_3	I think the supporting video tutorials were a useful complement to VR activities	
[EFF] Perceived learning effectiveness	EFF_1	I think the VR activities involving boxes are useful for improving comprehension of grammatical concepts	Taçgın (2020) Cronbach alpha = 0.796
	EFF_2	I think the VR activities involving plates are useful for improving memorization of new words	
	EFF_3	I think the textual/audio feedback received during VR activities is useful for improving comprehension of grammatical concepts	
[ENJ] Enjoyment	ENJ_1	The VR learning activities involving boxes were fun	Lin et Al. (2002) Cronbach alpha = 0.713
	ENJ_2	The VR learning activities involving plates were fun	
	ENJ_3	In the near future, I would like to engage in VR learning activities again, exploring new topics	
[CON] Comfort	CON_1	I quickly understood how to move within the VR activities.VR	Brunnström et al. (2020) Cronbach alpha = 0.697
	CON_2	The VR activities were comfortable (e.g., I did not experience physical discomfort such as eye strain, etc.)	
	CON_3	The physical space in which I moved while carrying out the VR activities was wide enough to have a comfortable experience	
	CON_4	Overall, I am satisfied with the learning content of VR activities	

All items were administered in Italian language using a seven-point Likert scale and have been translated in this table.

Author contributions CRediT authorship contribution statementMario Vallarino: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Software, Visualization, Writing– original draft, Writing– review and editing. Rita Cersosimo: Writing– original draft (Sect. 2.3), Writing– review and editing. Ilaria Torre: Methodology, Supervision, Writing– review and editing, Funding acquisition. Gianni Vercelli: Conceptualization, Methodology, Supervision, Project administration, Writing– review and editing.

Funding Open access funding provided by Università degli Studi di Genova within the CRUI-CARE Agreement. This work was partially supported by the PRIN 2022 PNRR Project (grant number P20227PEPK), funded by the European Union– Next Generation EU.

Declarations

Informed consent All participants provided informed consent upon their voluntary enrollment in the experiment.

References

Al-Azawei A, Baiee WR, Mohammed MA (2019) Learners' experience towards e-assessment tools: a comparative study on virtual reality and moodle quiz. Int J Emerg Technol Learn (Online) 14:34

Alfadil M (2020) Effectiveness of virtual reality game in foreign language vocabulary acquisition. Comput Educ 153:103893
Andrews M, Vigliocco G, Vinson D (2009) Integrating experiential and distributional data to learn semantic representations. Psychol Rev 116:463
Asher JJ (1966) The learning strategy of the total physical response: a review. Mod Lang J 50:79–84
Asher, J. J. (1977). *Learning Another Language through Actions: The Complete Teacher's Guidebook.* Los Gatos, Calif.: Sky Oaks Productions.
Aziz-Zadeh L, Damasio A (2008) Embodied semantics for actions: findings from functional brain imaging. J Physiol-Paris 102:35–39
Bailey B, Bryant L, Hemsley B (2022) Virtual reality and augmented reality for children, adolescents, and adults with communication disability and neurodevelopmental disorders: a systematic review. Rev J Autism Dev Disord 9:160–183
Barsalou LW (2008) Grounded cognition. Annu Rev Psychol 59:617–645
Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C et al (2021) Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. JMIR Serious Games 9:e29080
Biber D, Nekrasova T, Horn B (2011) The effectiveness of feedback for L1-English and L2-writing development: A meta-analysis. ETS Research Report Series 2011:i–99
Bloom, BS (1968). Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. Evaluation comment, 1, n2.
Bonato F, Bubka A, Palmisano S (2009) Combined pitch and roll and cybersickness in a virtual environment. Aviat Space Environ Med 80:941–945
Boulenger V, Hauk O, Pulvermüller F (2009) Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. Cereb Cortex 19:1905–1914
Brunnström K, Dima E, Qureshi T, Johanson M, Andersson M, Sjöström M (2020) Latency impact on quality of experience in a virtual reality simulator for remote control of machines. Signal Process Image Commun 89:116005
Chang E, Kim HT, Yoo B (2020) Virtual reality sickness: a review of causes and measurements. Int J Human-Comput Int 36:1658–1682
Conde Gafaro, B. (2019). Exploring self-regulated language learning with MOOCs. Journal of Interactive Media in Education, 2019.
Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. https://doi.org/10.2307/249008
Della Putta P, Suñer F (2023) Using the body to activate the brain. Research trends and issues. Rev Cogn Linguist 21:1–8
Desai RH, Conant LL, Binder JR, Park H, Seidenberg MS (2013) A piece of the action: modulation of sensory-motor regions by action idioms and metaphors. Neuroimage 83:862–869
Dhimolea TK, Kaplan-Rakowski R, Lin L (2022) A systematic review of research on high-immersion virtual reality for language learning. TechTrends 66:810–824
Edge D, Cheng, KY, Whitney M (2013). SpatialEase: learning language through body motion. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (pp. 469–472).
Elliman J, Loizou M, Loizides F (2016). Virtual reality simulation training for student nurse education. 2016 8th international conference on games and virtual worlds for serious applications (VS-games), (pp. 1–2).
Ellis R (2021) Task-based language teaching. Springer
Fast-Berglund Å, Gong L, Li D (2018) Testing and validating extended reality (xR) technologies in manufacturing. Procedia Manuf 25:31–38
Fein AD (2017). Multimedia learning: principles of learning and instructional improvement in Massive, Open, Online Courses (MOOCS).
Ferdinand J, Soller S, Hahn JU, Parong J, Göllner R. (2023). Enhancing the effectiveness of virtual reality in science education through an experimental intervention involving students' perceived usefulness of virtual reality.
Fernandes AS, Feiner SK (2016). Combating VR sickness through subtle dynamic field-of-view modification. 2016 IEEE symposium on 3D user interfaces (3DUI), (pp. 201–210).
Fowler C (2015) Virtual reality and learning: where is the pedagogy? Br J Educ Technol 46:412–422
Freina L, Ott M (2015) A literature review on immersive virtual reality in education: state of the art and perspectives. Int Sci Conf Elearn Softw Educ 1:10–1007
Garcia S, Kauer R, Laesker D, Nguyen J, Andujar M (2019). A virtual reality experience for learning languages. Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, (pp. 1–4).
Glenberg AM, Gutierrez T, Levin JR, Japuntich S, Kaschak MP (2004) Activity and imagined activity can enhance young children's reading comprehension. J Educ Psychol 96:424
Glenberg AM, Goldberg AB, Zhu X (2011) Improving early reading comprehension using embodied CAI. Instr Sci 39:27–39
Gruber A, Wagner M (2024) Fostering interdisciplinary intercultural citizenship with virtual reality in a world language virtual exchange project. Intercult Commun Educ 7:1304–1304
Huang CL, Luo YF, Yang SC, Lu CM, Chen A-S (2020) Influence of students' learning style, sense of presence, and cognitive load on learning outcomes in an immersive virtual reality learning environment. J Educ Comput Res 58:596–615
Huang HM, Liaw SS (2018). An analysis of learners' intentions toward virtual reality learning based on constructivist and technology acceptance approaches. International Review of Research in Open and Distributed Learning, 19.
Jamieson GH (1970) Transfer of learning under two conditions of instruction: programmed and guided discovery. Program Learn Educ Technol 7:113–119
Jiang H, Vimalesvaran S, Wang JK, Lim KB, Mogali SR, Car LT (2022) Virtual reality in medical students' education: scoping review. JMIR Med Educ 8:e34860
Jirak D, Menz MM, Buccino G, Borghi AM, Binkofski F (2010) Grasping language–a short story on embodiment. Conscious Cogn 19:711–720
Jitpaisarnwattana N, Reinders H, Darasawang P (2019) Language MOOCs: an expanding field. Technol Lang Teach Learn 1:21–32
Jo H, Park D-H (2023) Affordance, usefulness, enjoyment, and aesthetics in sustaining virtual reality engagement. Sci Rep 13:15097
Jusslin S, Korpinen K, Lilja N, Martin R, Lehtinen-Schnabel J, Anttila E (2022) Embodied learning and teaching approaches in language education: a mixed studies review. Educ Res Rev 37:100480
Lacey S, Stilla R, Sathian K (2012) Metaphorically feeling: comprehending textural metaphors activates somatosensory cortex. Brain Lang 120:416–421
Lee J, Kim J, Choi JY (2019) The adoption of virtual reality devices: the technology acceptance model integrating enjoyment, social interaction, and strength of the social ties. Telemat Informatics 39:37–48
Legault J, Zhao J, Chi Y-A, Chen W, Klippel A, Li P (2019) Immersive virtual reality as an effective tool for second language vocabulary learning. Languages (Basel) 4:13
Lin JJW, Duh HBL, Parker DE, Abi-Rached H, Furness TA (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. Proceedings ieee virtual reality 2002, (pp. 164–171).

Lindgren R, Johnson-Glenberg M (2013) Emboldened by embodiment: six precepts for research on embodied learning and mixed reality. Educ Res 42:445–452

Mahon BZ, Caramazza A (2008) A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. J Physiol-Paris 102:59–70

Makowski D, Sperduti M, Nicolas S, Piolino P (2017) “Being there” and remembering it: Presence improves memory encoding. Conscious Cogn 53:194–202

Makransky G, Mayer RE (2022) Benefits of taking a virtual field trip in immersive virtual reality: evidence for the immersion principle in multimedia learning. Educ Psychol Rev 34:1771–1798

Makransky G, Terkildsen TS, Mayer RE (2019) Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learn Instr 60:225–236

Mayer RE (2005) The Cambridge handbook of multimedia learning. Cambridge University Press

Mayer RE (2008) Applying the science of learning: evidence-based principles for the design of multimedia instruction. Am Psychol 63:760

Mayer RE (2010) Applying the science of learning to medical education. Med Educ 44:543–549

Mayer RE (2014) Incorporating motivation into multimedia learning. Learn Instr 29:171–173

Mayer RE (2020) Multimedia Learning. Cambridge University Press

Mayer RE, Anderson RB (1991) Animations need narrations: an experimental test of a dual-coding hypothesis J. Educ Psychol 83:484

Mayer RE, Anderson RB (1992) The instructive animation: helping students build connections between words and pictures in multimedia learning. J Educ Psychol 84:444

Mayer RE, Gallini JK (1990) When is an illustration worth ten thousand words? J Educ Psychol 82:715

Mayer RE, Bove W, Bryman A, Mars R, Tapangco L (1996) When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. J Educ Psychol 88:64

Milgram P, Takemura H, Utsumi A, Kishino F (1995) Augmented reality: A class of displays on the reality-virtuality continuum. Telemanipulator and Telepresence Technologies 2351:282–292

Montoya-Rodríguez MM, de Souza Franco V, Tomas Llerena C, Molina Cobos FJ, Pizzarossa S, García AC, Martínez-Valderrey V (2023). Virtual reality and augmented reality as strategies for teaching social skills to individuals with intellectual disability: A systematic review. Journal of Intellectual Disabilities, 27, 1062–1084.

Panigagua A, Istance D (2018). Teachers as Designers of Learning Environments: The Importance of Innovative edagogies. Educational Research and Innovation, Paris, OECD Publishing

Parmaxi A (2023) Virtual reality in language learning: a systematic review and implications for research and practice. Interact Learn Environ 31:172–184

Pekrun R, Goetz T, Frenzel AC, Barchfeld P, Perry RP (2011) Measuring emotions in students’ learning and performance: the achievement emotions questionnaire (AEQ). Contemp Educ Psychol 36:36–48

Petersen GB, Petkakis G, Makransky G (2022) A study of how immersion and interactivity drive VR learning. Comput Educ 179:104429

Pinto, D., Peixoto, B., Krassmann, A., Melo, M., Cabral, L., & Bessa, M. (2019). Virtual reality in education: Learning a foreign language. *New Knowledge in Information Systems and Technologies: Volume 3*, (pp. 589–597).

Plotzky C, Lindwedel U, Sorber M, Loessl B, König P, Kunze C, Kugler C, Meng M (2021) Virtual reality simulations in nurse education: a systematic mapping review. Nurse Educ Today 101:104868

Ponce HR, Mayer RE (2014) An eye movement analysis of highlighting and graphic organizer study aids for learning from expository text. Comput Human Behav 41:21–32

Pouke M, Tiiro A, LaValle SM, Ojala T (2018). Effects of visual realism and moving detail on cybersickness. *2018* IEEE Conference on Virtual Reality and 3D User Interfaces (VR), (pp. 665–666).

Ricci G, De Crescenzio F, Santhosh S, Magosso E, Ursino M (2022) Relationship between electroencephalographic data and comfort perception captured in a virtual reality design environment of an aircraft cabin. Sci Rep 12:10938

Risi D, Palmisano S (2019) Effects of postural stability, active control, exposure duration and repeated exposures on HMD induced cybersickness. Displays 60:9–17

Sagnier C, Loup-Escande E, Lourdeaux D, Thouvenin I, Valléry G (2020) User acceptance of virtual reality: an extended technology acceptance model. International Journal of Human-Computer Interaction 36:993–1007

Savignon, S. J. (1972). Communicative Competence: An Experiment in Foreign-Language Teaching. Language and the Teacher: A Series in Applied Linguistics, Volume 12.

Shah, D. (2023). *Massive List of MOOC Platforms Around the World in 2023*. Retrieved from Class Central: https://www.classcentral.com/report/mooc-platforms/

De Simone, F., Li, J., Debarba, H. G., El Ali, A., Gunkel, S. N., & Cesar, P. (2019). Watching videos together in social virtual reality: an experimental study on user’s QoE. 2019 IEEE Conference on virtual reality and 3d user interfaces (VR), (pp. 890–891).

Soliman M, Pesyridis A, Dalaymani-Zad D, Gronfula M, Kourmpetis M (2021) The application of virtual reality in engineering education. Appl Sci 11:2879

Taçgın Z (2020) The perceived effectiveness regarding immersive virtual reality learning environments changes by the prior knowledge of learners. Educ Inf Technol 25:2791–2809

Tang S (2017) Learning mechanism and function characteristics of MOOC in the process of higher education. Eurasia J Math Sci Technol Educ 13:8067–8072

Tellier M (2008) The effect of gestures on second language memorisation by young children. Gesture 8:219–235

Vallarino, M., & Vercelli, G. (2023). A proposal for a virtual reality method in language learninG. EDULEARN23 Proceedings, (pp. 596–600).

Vázquez, C., Xia, L., Aikawa, T., & Maes, P. (2018). Words in motion: Kinesthetic language learning in virtual reality. 2018 IEEE 18th International Conference on advanced learning technologies (ICALT), (pp. 272–276).

Wheeler SG, Engelbrecht H, Hoermann S (2021) Human factors research in immersive virtual reality firefighter training: a systematic review. Front Virtual Real 2:671664

Willems RM, Casasanto D (2011) Flexibility in embodied language understanding. Front Psychol 2:116

Wilson NL, Gibbs RW Jr (2007) Real and imagined body movement primes metaphor comprehension. Cogn Sci 31:721–731

Winstone NE, Boud D (2022) The need to disentangle assessment and feedback in higher education. Stud High Educ 47:656–667

Wu Y, Chen S-C, Lin I-C (2019) Elucidating the impact of critical determinants on purchase decision in virtual reality products by analytic hierarchy process approach. Virtual Reality 23:187–195

Xie B, Liu H, Alghofaili R, Zhang Y, Jiang Y, Lobo FD, Li C, Li W, Huang H, Akdere M, Mousas C, Yu L-F (2021) A review on virtual reality skill training applications. Front Virtual Real 2:645153

Yudintseva A (2024) An exploration of low-and high-immersive virtual reality modalities for willingness to communicate in English as a second language. Computers & Education: X Reality 5:100076

Zhang X, Jiang S, Ordóñez de Pablos P, Lytras MD, Sun Y (2017) How virtual reality affects perceived learning effectiveness:

a task–technology fit perspective. Behav Inform Technol 36:548–556

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