Diverse approaches to learning with immersive Virtual Reality identified from a systematic review

Mihye Won a,*, Dewi Ayu Kencana Ungu a, Henry Matovu a, David F. Treagust a, Chin-Chung Tsai b, Jungho Park c, Mauro Mocerino d, Roy Tasker e

a School of Education, Curtin University, Australia
b Program of Learning Sciences & Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taiwan
c Lincoln Agritech, New Zealand
d School of Molecular and Life Sciences, Curtin University, Australia
e School of Science, Western Sydney University, Australia

ARTICLE INFO

Keywords:
Immersive virtual reality
Systematic literature review
Learning tasks
Learning design
Cluster analysis

ABSTRACT

To investigate how learning in immersive Virtual Reality was designed in contemporary educational studies, this systematic literature review identified nine design features and analysed 219 empirical studies on the designs of learning activities with immersive Virtual Reality. Overall, the technological features for physical presence were more readily implemented and investigated than pedagogical features for learning engagement. Further analysis with k-means clustering revealed five approaches with varying levels of interactivity and openness in learning tasks, from watching virtual worlds passively to responding to personalised prompts. Such differences in the design appeared to stem from different practical and educational priorities, such as accessibility, interactivity, and engagement. This review highlights the diversity in the learning task designs in immersive Virtual Reality and illustrates how researchers are navigating practical and educational concerns. We recommend future empirical studies recognise the different approaches and priorities when designing and evaluating learning with immersive Virtual Reality. We also recommend that future systematic reviews investigate immersive Virtual Reality-based learning not only by learning topics or learner demographics, but also by task designs and learning experiences.

1. Introduction

1.1. Immersive Virtual Reality

Educational researchers have explored the possibilities of various tools and resources to provide improved learning experiences (Dede, Jacobson, & Richards, 2017). One of the latest technological innovations—immersive Virtual Reality (IVR)—has become an area of interest amongst educators and researchers (Sherman & Craig, 2018). Different from other Virtual Reality platforms (e.g., second life or massive online games on 2D screens), IVR uses precise, real-time motion tracking devices and powerful graphic processors to render continuous stereoscopic images on head-mounted display (HMD). When combined with absorbing narratives or
convincing contexts, IVR technology can provide an illusion that users are interacting with 3D objects in a virtual environment as they would in the real world (Bailenson, 2018; Feng, González, Amor, Lovreglio, & Cabrera-Guerrero, 2018). Consequently, IVR has been hailed as a powerful 3D visualisation tool that could change the way people perceive and experience virtual environments (Slater, 2017). With increasing accessibility and representational fidelity of the IVR technology, the number of researchers investigating the educational possibilities of IVR technology is dramatically increasing (Radianti, Majchrzak, Fromm, & Wohlgenannt, 2020; Wu et al., 2020a).

1.2. What we know and do not know on IVR

Previous systematic literature review studies reported that research participants appreciated IVR-based activities as an engaging novel learning experience and expressed their willingness to go through similar IVR-based learning activities in the future (Di Natale, Repetto, Riva, & Villani, 2020; Jensen & Konradsen, 2018; Radianti et al., 2020). Previous meta-analysis studies revealed that IVR had a moderately positive effect on learning compared to lectures or other less immersive media (Hamilton, McKechnie, Edgerton, & Wilson, 2021; Wu et al., 2020a). However, there are conflicting accounts of in what circumstances IVR is more effective. Jensen and Konradsen’s (2018) systematic literature review concluded that IVR was moderately useful for skills acquisition but not for other learning objectives. Hamilton et al.’s (2021) systematic literature review indicated that half of the studies on cognitive skills and on knowledge acquisition had positive effects, but that the rest of the studies (e.g., on skill acquisition) demonstrated no significant effects on learning. Wu, Yu, and Gu (2020) meta-analysis showed that IVR was effective for younger learners, in the form of simulations or virtual world representations, and for science education and development of specific skills. Coban et al.’s (2022) meta-analysis reported that IVR was effective for K-12 education and for architecture and engineering, but not for science or surgical training. With conflicting accounts of ‘what worked’ in relation to different learning objectives or discipline areas, it is difficult to understand the status of educational IVR studies and use the information to further investigate educational possibilities of IVR. A different approach to review and synthesize the existing literature on educational IVR could help.

1.3. What we need to know

When investigating the educational possibilities of technology-supported learning environments, researchers recommend identifying the unique characteristics of the environments (i.e., what makes IVR compelling) and investigating how those characteristics interact with educational contexts and pedagogical approaches (Dalgarno & Lee, 2010; Klopfer, 2008; Mikropoulos & Natsis, 2011). However, existing literature reviews on IVR as a particular type of technology have not systematically documented how researchers understood the unique characteristics of IVR and how they approached IVR-based learning. What do researchers see as the strengths of IVR and IVR-based learning? How do they capitalise the advantages of IVR to support learning in virtual environments? Documenting how educational researchers designed learning activities using the characteristic features of IVR would provide an insight to the educational possibilities of IVR, as Wu, Lee, Chang, and Liang (2013) did for augmented reality and Abdul Jabbar and Felicia (2015) for game-based learning. Unfortunately, there is no systematic literature review that methodically charted different designs of learning activities in relation to the integration of IVR’s key characteristics or identified common design approaches. Radianti et al.’s (2020) literature review observed several design elements, but did not delineate the levels of integrations and did not demonstrate connections to specific learning tasks, educational contexts, or pedagogical approaches.

1.4. Rationale of this review

As a group of multi-disciplinary educational researchers, the authors of this review considered an alternative way to examine the design of existing IVR studies so that it would effectively guide educational researchers like ourselves in designing IVR applications to investigate the educational possibilities of IVR. It is not the intention of this review to evaluate the quality of IVR applications or the rigor of educational research studies. Rather, by documenting how IVR-based learning activities have been constructed and executed for different learning objectives and how they contributed to creating unique learning experiences, this systematic literature review aims to recommend a concrete way forward for educational researchers to investigate the educational possibilities of IVR.

2. Conceptual framework: design features for educational IVR

To examine the design of IVR-based learning environments in the contemporary literature, this review first identified the key design features which would contribute to productive educational experiences in IVR environments. In investigating educational possibilities of IVR-based learning environments, both technological characteristics and pedagogical aspects need to be considered. This review adapts Dede’s (2009; 2017) immersive interface design, which includes sensory, actional, narrative, and social features. Dede’s immersive interface design framework offers a comprehensive list of ways that people can utilise the IVR with a scope for expansion and elaboration.

IVR application designers might first consider the technological capabilities of IVR and integrate representationally authentic sensory stimuli to induce the feeling that virtual objects and environments are real (sensory—representational fidelity). Establishing a robust IVR interface design is also an important technological design decision because it would allow users to act naturally and intuitively in virtual worlds as if they are interacting in real environments (actional—interactions). From a pedagogical point of view, the way in which the learning content is woven in IVR applications is a critical consideration to engage users emotionally and
intellectually in completing the tasks (narrative—task content). Another key pedagogical decision is how to support learning as learners interact in virtual environments (social—constructive support). Within the premise of these four key design features (two technological and two pedagogical), the authors of this review identified specific design features for IVR which are elaborated below. A concise description of each design feature is shown in Table 1.

2.1. Sensory

The perception of being in a virtual environment, called physical presence (Lee, 2004), can be induced by various sensory-motor stimuli (Dede, 2009; Dede et al., 2017). For example, when users have a realistic panoramic 3D view of numerous skyscrapers laying beneath their feet in a virtual world, they would perceive that they are transported to the top of the tallest skyscraper and they would be reluctant to take a step forward, afraid that they would fall into an abyss even though they ‘know’ they are safe to walk around in the physical world. As users experience the computer-generated authentic 3D graphics and other multiple sensory-motor cues, their sense of ‘what is real’ and ‘where they are’ becomes distorted (Slater, 2017).

Researchers (e.g., Dalgarno & Lee, 2010; Sanchez-Vives & Slater, 2005) have identified what sensory stimuli contribute to inducing physical presence, to include visual (3D graphics), audio (spatial sound effects), and haptic feedback. Among them, visual, the advanced 3D visualisation of IVR technology, plays a key role in creating physical presence. Different IVR hardware offer varying levels of representational fidelity of the 3D graphics, and the selection of the IVR hardware (e.g., stereoscopic display and wider fields of view) as well as the software architecture influence the level of physical presence in users (Cummings & Bailenson, 2015).

Audio or sound effects could also increase the level of physical presence (Sanchez-Vives & Slater, 2005). For example, instead of hearing people’s chatter in the physical world, hearing rustling of leaves and roaring of a wild cat in a jungle (as in Google Expeditions©) could give a sense of being transported to a virtual world. The audio could be used to induce emotional involvement in IVR tasks as well as offering subtle cues. For example, in Beat Game’s Beat Saber©, the background music with upbeat rhythms gives users hints when they need to swing their arms to hit the virtual objects, and it also uplifts the mood of the users to enjoy the IVR environment.

Haptic responses involving texture and force feedback of an object provides a sense of ‘being in a new environment’ (Sanchez-Vives & Slater, 2005). If users hit a concrete wall in a virtual environment and feel the physical wall with their hands at the same time, they perceive virtual and physical worlds merging. Some IVR theatres create a hybrid space where the physical arrangement of a room mirrors a virtual environment (e.g., driver seat, steering wheel, and brake pedal) to provide adequate haptic responses.

2.2. Actional

The interactions in a virtual environment could enhance the feeling of being and acting in a virtual world (Dede, 2009; Dede et al., 2017). For example, in an interactive archery simulation (Valve’s Longbow©), users can pick up a bow and arrow, draw the bow, and aim it at moving targets. In a similar fashion to the physical world, the bow must be drawn each time along a realistic draw path, the draw length determines the speed of the arrow, and the user must aim the bow and arrow well. Upon release, the arrow then flies as though it is affected by gravity. By repeating the movement and seeing its consequences in real time, users come to believe that they are actually shooting arrows to a target rather than holding IVR controllers.

When the application is responsive and well designed to allow intuitive and natural actions in a virtual world, users forget they are

<table>
<thead>
<tr>
<th>Design features</th>
<th>Descriptions</th>
<th>Subcategories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensory</td>
<td>Representational fidelity. The presented virtual environment is representationally sound for learners to feel that the virtual objects and places are authentic or real.</td>
<td>• Visual: High quality graphics with motion tracking • Audio: Sound effects • Haptic and others: Tactile, force feedback and appeals to other senses • Interactivity: Computer-user interactions and levels of control • Movements: Relevant physical body movements to experience and complete a task • Roles: The role of the user and the consequentiality of their actions • Contexts: Relatable storyline in relevant contexts • Challenges: Opportunities to apply themselves and learn from the experience • Social interactions: Mediated social interactions</td>
</tr>
<tr>
<td>Actional</td>
<td>Intuitive interface design. The actions in a virtual environment feel natural and intuitive for learners to feel they are making real changes in the environment.</td>
<td></td>
</tr>
<tr>
<td>Narrative</td>
<td>Engaging content and task. The content and tasks are relevant and meaningful for learners to feel emotionally and intellectually engaged.</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>Constructive support. The learners and learning are supported through social interactions.</td>
<td></td>
</tr>
</tbody>
</table>
interacting in a made-up world and become involved in their interactions in the virtual world (Dalgarno & Lee, 2010). Two highly related design aspects are involved in this regard: interactivity and embodied movement. Interactivity or interactive interface design allows a high level of user control and produces responsive output (Roussou, Oliver, & Slater, 2006). Natural interactions occur when relevant body movements are recognised as user inputs with real-time motion tracking (Sanchez-Vives & Slater, 2005; Slater, 2017). Instead of letting users select an option of a projectile motion from a drop-down menu to watch an arrow fly as in a movie, an interactive, embodied IVR application would allow users to pick up a bow and arrows with their hands and shoot the arrows themselves as they would in the physical world.

### 2.3. Narrative

The content and the nature of the tasks play a critical part in drawing users’ attention and engaging them in meaningful learning (Barab, Gresalfi, & Ingram-Goble, 2010). The art of organising the content for increased participation in learning is difficult to pin down, but the following components are often considered in the context of educational computer applications: roles, contexts, and challenges (Abdul Jabbar & Felicia, 2015; Barab et al., 2010).

Learners feel agency to complete a set of tasks when they are given a clear role to carry out consequential tasks. In IVR environments, virtual body ownership—one of the most distinguishing and intriguing features of IVR—facilitates learners to become someone else and draw them deeply into the made-up world (Slater, 2017; Yee & Bailenson, 2007). When designed well with plausible scenarios and comprehensive movement tracking, learners could see their body movements mirrored in the form of avatars, to assume a new role readily and experience the consequences of their actions as real. This feature has been reported to be successful in helping users question their assumptions and change behaviours (Jacobson, 2017; Slater, 2017; Yee & Bailenson, 2007).

Skillfully crafted storylines in relevant contexts are effective at engaging learners in learning (Abdul Jabbar & Felicia, 2015). For example, a charming fantasy VR film, Allumette©, compels users to be emotionally involved and empathise with the protagonist, a young orphan girl who endures a tragedy and finds kindness and hope. When educators aim for intellectual involvement as well as emotional involvement through educational applications, it is critical to integrate engaging and meaningful storylines to serve the purpose of contextualising learning and motivating learners to expand their knowledge, skills and attitudes (Barab et al., 2010).

Challenges and achievements are also an important aspect of task design. The tasks need to be clear and pose just the right level of challenge so that learners can immerse themselves to complete the task and feel the sense of achievement (Csikszentmihalyi, 2014). To match learners’ knowledge and skills, IVR applications could offer multiple levels of challenges and multiple attempts.

### 2.4. Social

Constructive social interactions in a virtual environment enable learners to feel part of a virtual community and engage in fruitful exchange of ideas (Ode, 2009). Learners may share a virtual space with other people in IVR or perceive virtual actors, such as intelligent pedagogical agents, as actual social actors. Either way, productive social interactions are critical not only to share and reflect on learners’ own ideas but also to offer and receive feedback on their performance in virtual environments to enhance the quality of experience and knowledge construction (Dalgarno & Lee, 2010; Mikropoulos & Natsis, 2011). Researchers observed that collaborative interactions in technology supported learning environments not only deepen the feeling of enjoyment and engagement in learning, but also help develop students’ knowledge, communication skills, higher-order thinking skills, and problem solving skills (Chen, Wang, Kirschner, & Tsai, 2018; Garzón, Kinshuk, et al., 2020; Klopfner, 2008). With the development of artificial intelligence in recent years, the role of intelligent pedagogical agents could become important to provide just-in-time feedback to learners and work as a friendly companion or tutor (Soliman & Guehl, 2010, pp. 24–28).

Based on the key design features identified above, this review investigated the following research questions:

- How are technological and pedagogical features implemented in educational IVR studies?
- How do the implemented features relate to the learning tasks in IVR?
- What approaches do researchers take when designing IVR-based learning?

### 3. Method

#### 3.1. Search and selection of empirical studies

To understand which immersive learning components have been adopted in educational IVR application development and what researchers have found as educational affordances, four authors of this review (Au1, Au2, Au3, & Au6) searched and screened eligible educational empirical studies following the updated preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines (Page et al., 2021). Scopus, ProQuest, Web of Science, and Google Scholar were identified as the main databases. Scopus, ProQuest, and Web of Science are well-known repositories of scholarly articles in a range of disciplinary areas. Google Scholar offers a vast range of articles, including conference proceedings and book chapters.

Using the following search string, two authors (Au1 & Au6) searched the three databases (Scopus, ProQuest, and Web of Science) to include educational studies with original empirical data using immersive VR devices.
“immersive virtual reality”
AND (education OR learning OR teaching)
AND (participant OR student OR learner)
AND (HMD OR headset)

As a nested search string could not be established for Google Scholar, the authors (Au1 & Au6) used a slightly modified search string. Using a near-term search strategy, the highest hit word was chosen amongst each group of words—learning from the group of (education, learning, and teaching), and participant from the group of (participant, student, and learner). Because HMD (head-mounted display) and headset were specific words with less overlap, two separate searches were conducted and later combined.

“immersive virtual reality” learning participant HMD
“immersive virtual reality” learning participant headset

The search field was set for all text in the document. Where possible, the document types were limited to full-text and peer reviewed articles, written in English. The coverage dates were from 2016 till 2021. As Google Search Trends demonstrates, the interest in VR dramatically increased in 2016 with the release of much anticipated IVR headsets (HTC VIVE and Oculus Rift CV), prompting a higher number of research studies with IVR headsets (Sherman & Craig, 2018).

The database search yielded 15,421 records (Scopus 786, ProQuest 747, Web of Science 38, and Google Scholar 6,940 for HMD and 6,910 for headset). Because Google Scholar displayed the first 1,000 records for each search query, the search was done for each year to retain most relevant studies and reduce data loss. Because the number of records was unmanageably high, an additional screening process was brought in. The number of citations is often used as a measure of scientific impact and relevance in academia (Aksnes, Langfeldt, & Wouters, 2019). Using a Python package scholarly.py, the total number of citations of each study was extracted through Google search with the exact title of the study with quotation marks on February 16, 2022. For web scraping, two proxy application programming interface (API) services, SerpAPI and ScraperAPI, were used. Assuming the citation numbers would increase linearly over time, increasing citation numbers deemed appropriate as the filter. To select roughly 20% of relevant and impactful studies published in each year, 5 citations were set as the bar for the studies published in 2021, with the increment of 5 for each passing year. The scripts for the initial screening process and for an XML file generation are included as online supplementary material. After converting the XML file to an EndNote library, the same two authors (Au1 & Au6) manually removed the studies with less than 5 citations per year to yield 2,785 studies. After removing duplicates (1,215), total of 1,728 studies remained (136 studies from 2016, 218 studies from 2017, 270 studies from 2018, 418 studies from 2019, 426 studies from 2020, to 260 studies from 2021).

Two other authors (Au2 & Au3) read through the abstracts and the document information of 1,728 studies for further screening. The eligibility criteria were that the study (a) described educational use of fully immersive VR with a headset; (b) was conducted in the field of education; (c) presented an original empirical data from participants; (d) was published in peer-reviewed journals, books, or reputable conference proceedings listed in SCImago; and (e) provided full text available in English. This screening process removed

Fig. 1. Flow chart of the literature identification and screening process.
studies on mixed or augmented reality devices (e.g., Zhou, Segura, Duval, John, & Isbister, 2019), studies in the fields of therapy, rehabilitation, or business (e.g., Gold & Mahar, 2017; Tussyadiah, Wang, Jung, & tom Dieck, 2018), and theoretical or secondary data analysis studies (e.g., Parsons, 2016), and application development reports without empirical findings (e.g., Pinter et al., 2020). After the screening, 276 studies remained.

The two authors (Au2 & Au3) read through the full text applying the criteria, further removing studies based on the same inclusion criteria. Two more criteria were added at this stage: (f) provided sufficient information on the IVR application features, learning tasks, and evaluation of learning; and (g) reported original data that are significantly different from other included studies. After looking through electronic supplementary materials and searching for additional information on the internet, the authors removed 25 studies due to there being insufficient information (e.g., Hadley et al., 2018; Ochs et al., 2019). Five studies deemed similar to already included studies in terms of authors, IVR applications, participant groups, and educational contexts were removed. Subsequently, a total of 219

Table 2
Categorisation of the integration of design features in educational IVR applications.

<table>
<thead>
<tr>
<th>Design features</th>
<th>Integration levels</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensory (representational fidelity of virtual environment)</td>
<td>Visual</td>
<td>Low resolution with a lower refresh rate and a smaller field of view to offer a limited sense of physical presence (e.g., Google Cardboard or Samsung Gear as in Lin &amp; Wang, 2021)</td>
<td>Medium resolution graphics, refresh rate and field of view (e.g., Oculus Go or Rift DK2 as in Ding, Brinkman, &amp; Neerincx, 2020)</td>
<td>High resolution graphics with a high refresh rate and a wider field of view to heighten the sense of physical presence (e.g., HTC Vive Pro or Oculus Rift S as in Ropelato, Zünd, Magnenat, Menozzi, &amp; Sumner, 2018)</td>
</tr>
<tr>
<td></td>
<td>Audio</td>
<td>Minimal audio and sound effects</td>
<td>Audio instructions with minimal sound effects or arrangement (e.g., audio instruction as in Gerry, 2017)</td>
<td>Effective immersive sound effects for location and direction (e.g., realistic spatial sound effects and blocking the sound of physical environment as in Calver &amp; Abadia, 2020)</td>
</tr>
<tr>
<td></td>
<td>Haptic</td>
<td>No haptic or tactile responses</td>
<td>Some haptic feedback (e.g., controller vibration as in Lohre, Bois, Athwal, &amp; Goel, 2020)</td>
<td>Realistic force feedback and additional sensory responses through a hybrid setup (e.g., holding a surgical instrument with variable resistance as in Burre et al., 2019)</td>
</tr>
<tr>
<td>Actional (interface design for natural and intuitive actions)</td>
<td>Interactivity</td>
<td>Minimal interactivity and user control (e.g., passively watching a 3D video as in Petersen, Klingenberg, Mayer, &amp; Makransky, 2020)</td>
<td>Medium interactivity and user control (e.g., selecting different options to proceed, as in Dunnagan et al., 2020)</td>
<td>High interactivity and intuitive user control to feel the interactions are natural and rule-bound (e.g., designing and building a wall by layering bricks within Minecraft VR as in Southgate et al., 2019)</td>
</tr>
<tr>
<td></td>
<td>Embodied movement</td>
<td>No obvious body movement expected, except head movement and button clicks for navigation in a virtual environment (e.g., no gesture or action involved as in Harrington et al., 2018)</td>
<td>Some body movement is expected to carry out the learning tasks (e.g., speaking in public in van Ginkel et al., 2019)</td>
<td>Whole body movement is essential to carry out the learning tasks (e.g., making a 3D design with a virtual brush of Google Tilt Brush as in Yang et al., 2018)</td>
</tr>
<tr>
<td>Narrative (content and task for engagement)</td>
<td>Roles</td>
<td>User’s presence not recognised or no specific roles given (e.g., spectator as in Chirico et al., 2020)</td>
<td>A specific role with specific tasks (mission) to complete (e.g., hospital visitor evacuating from an earthquake as in Lorreglio et al., 2018)</td>
<td>A clear role with a unique avatar to make consequential decisions (e.g., become a psychologist and a patient with full body avatars to confront their biases as in Slater, 2017)</td>
</tr>
<tr>
<td></td>
<td>Contexts and storylines</td>
<td>Generic context not aiming for deeper involvement (e.g., examining anatomical structures as in Moro, Stromberg, Raikos, &amp; Stirling, 2017)</td>
<td>Specific context for engagement but without a captivating storyline and characters (e.g., visiting the moon as in Kwon, 2019)</td>
<td>Skillfully crafted storyline in relevant context that appeals to learners’ experiences (e.g., experiencing homelessness as in Herrera, Bailenson, Weisz, Ogle, &amp; Zaki, 2018)</td>
</tr>
<tr>
<td></td>
<td>Challenges and achievement</td>
<td>Completion of a task as one-time experience (e.g., completing pre-surgical training once as in Pulijala et al., 2018a)</td>
<td>Offers further engagement and learning opportunities over time (e.g., experiencing three IVR simulations over time as in Pande et al., 2021)</td>
<td>Varying content and difficulty in tasks to optimise challenge levels and accommodate learners’ experiences (e.g., Ropelato et al., 2018) with opportunities to make mistakes and learn from them (e.g., Real et al., 2017)</td>
</tr>
<tr>
<td>Social (social support for learning)</td>
<td>Social interactions</td>
<td>No mediated social interactions for learning</td>
<td>Limited responses from peers, teachers, or pedagogic agents to support learning (e.g., pre-programmed limited responses as in Mcleery et al., 2020)</td>
<td>Extensive mediated social interactions for negotiated meaning making and to provide feedback on learning (e.g., peer-to-peer collaboration in a networked virtual environment as in Sasinka et al., 2019)</td>
</tr>
</tbody>
</table>
3.2. Categorisation of the design features of IVR applications

The authors established a set of categories to evaluate how each design feature was integrated to IVR applications in the selected empirical studies. An initial categorisation scheme was established prior to coding (Au1). Initial coding of selected studies led to extensive discussions among the authors (Au1, Au2, & Au3) regarding fair evaluation and differentiation amongst individual studies. Using a modified categorisation scheme, two authors (Au2 & Au3) read through each empirical study and categorised as low, medium, and high levels of integration of design features. The descriptors and example studies are included in Table 2. For statistical analysis, low level integration was recorded as 1, medium 2, and high 3. Two authors individually read and categorised the studies in their entirety (N = 219). The overall interrater reliability (Spearman’s correlation coefficient rho) was 0.97** (p < 0.01). After discussion amongst the authors (Au1, Au2, & Au3), any disagreement was resolved, and some wording were slightly modified to reflect the agreed-upon categorisation scheme. The final categorisation and descriptors are illustrated in Table 2. The categorisation results are included in the online supplementary material.

3.3. Categorisation of the learning tasks

To characterise the learning activities in IVR studies in relation to the design features, this review adopted Merchant and colleagues’ (2014) categorisation of learning tasks in the context of three types of educational computer applications—virtual world representations, simulations, and games for learning. In the applications for virtual world representations, learners do what is impossible in the real world, such as observe and interact with imaginary objects, have a tour of made-up worlds, become someone else, or design new objects or environments. In simulations that imitate real events or situations, learners practice a series of tasks, learn how to respond to various scenarios, or test hypotheses. Games for learning adopt various game elements to absorb users’ attention to help achieve defined learning outcomes. The learning tasks may involve answering a series of questions with a scoring system (gamification) or involving physical/intellectual challenges (game play). The categorisation of learning tasks and example studies are included in Table 3.

In addition to the general learning tasks, the subject areas (e.g., medical, science, history, and design) and target audience (e.g., primary school students, secondary school students, university students, and adults) were also recorded to better understand the learning activities. A summary of each study with the identified tasks and contexts is included in the online supplementary material.

3.4. Analysis of the common design features in educational IVR studies

To understand how various design features were implemented in educational IVR studies, multiple statistical analyses were conducted. First, simple descriptive statistics were run with Microsoft Excel to provide an overall picture of how design features were integrated in educational IVR applications (Au1). Spearman’s correlation coefficient rho for ordinal data was calculated using a Python package, scipy.stats.

Second, the shared design characteristics within educational IVR studies were identified by adopting k-means clustering (Au6). As a popular unsupervised machine learning algorithm, k-means clustering aims to assign data into a set number of groups based on the similarities and differences of the data (Hearty, 2016). Scikit-learn (Pedregosa et al., 2011), a Python module in machine learning, was used for this operation. After experimenting with different numbers of clusters, the authors of this review chose five clusters because this number provided reliable and logical group separation. Regardless of the selection of initial centre locations of k-clusters, k-means
analysis converged the data into exactly the same five clusters with minimal errors after running 2,000 times with randomly selected five initial points, resembling the exact $k$-means algorithm (Wishart, 2005).

Third, based on the clustering information, the integration levels of design features were reconsidered to find common features that distinguished each cluster from the rest (Au1, Au2, & Au3). In this process, it was determined that the common features were linked to different learning tasks and thus design approaches. The common learning tasks and design features were used as the name of each approach. The studies in each cluster were reviewed again in relation to different learning activities (i.e., learning tasks and content areas) and pedagogical approaches.

4. Results

4.1. Integration of immersive learning design features and correlations

Educational IVR applications displayed diverse design features, but some design features were more readily adopted than others (see Fig. 2). The sensory—visual feature was the most readily adopted design feature with more than half of the applications recording high level of integration. Many applications were run in high-end IVR headsets to offer superior visual experience of 3D environments (126 out of 219 studies). Among the sensory features, haptic feedback was least adopted, having only 30 studies with some form of haptic responses. For audio, a huge number of applications incorporated audio instructions or narrations (170 studies), but realistic spatial sound effects were limited (13 studies).

For actional design considerations, three quarters of the applications integrated interactive interfaces (166 studies) going beyond passively watching 360° videos, and some of them allowed users to create virtual objects (10 studies). The integration of embodied movement was less common compared to interactivity. A little less than half of the studies (99 studies) required no body movement or simple point-and-click as ways to interact within virtual environments. Thirty applications were designed to encourage whole body movement to complete the tasks.

For narrative design considerations, three fifths of the studies assigned specific tasks and roles to learners (133 studies), but no study gave a unique full body avatar with kinematics to merge their roles and physical beings in virtual environments. Three fifths of the studies contextualised the tasks for learner engagement (134 studies), with a small number of studies integrating captivating storylines and personable virtual characters (8 studies). About two thirds of the studies adopted IVR applications as a short, one-time experience (140 studies) rather than expecting learners to progressively develop the knowledge and skills in IVR over time (79 studies).

For social design considerations, two thirds of the studies had no mediated social interactions for learning (147 studies). A third of the studies offered mediated social interactions in IVR applications in the form of interaction with semi-intelligent pedagogical agent or peer-to-peer interactions (72 studies).

Individual design features were compared to one another to find relations between them (see Table 4). Out of nine design features, visual, interactivity, embodied movement, and roles showed strong correlations with other design features. The strong correlation amongst three technological design features (visual, embodied movement and interactivity) indicates the link between interface design and hardware selection. As lower-end IVR headsets allow limited embodied movement, the studies with low visual tend to record low interactivity and embodied movement. When high-end headsets were employed, it was more likely to see interactive, movement-oriented features. Interestingly, embodied movement had a strong correlation with both interactivity and haptic feedback, but the correlation between interactivity and haptic feedback was weak. This implies that integration of physical body movement required both high interactivity and realistic haptic feedback, but haptic feedback was not necessary for highly interactive IVR applications.

![Fig. 2. Integration of design features (N = 219).](image-url)
Among pedagogical design features, roles showed a strong correlation with all other design features, except audio. It implies that researchers recognised assigning a specific role as important to help learners feel the ownership of the learning tasks and be engaged in learning. In contrast, two other narrative design features (i.e., contexts and challenges) and the social design feature showed moderate correlations with roles and with one another, but not so much with other technological design features. This observation implies that the pedagogical design features tended to be implemented together, irrespective of the implementation of technological design features.

4.2. Integration of design features in relation to learning tasks and contexts

Educational studies adopted IVR applications in the form of simulations (104 studies) or virtual world representations (85 studies) more than games for learning (30 studies). The studies in each application type provided relatively uniform rationales for adopting IVR, but the tasks and contexts varied.

IVR applications for virtual world representations were adopted to help learners visualise objects, events, or environments so that they experience something impossible or impractical in the real world and gain knowledge. Such applications allowed students to interact with 3D models of molecules, cells, planets, or human anatomy, to visualise and build scientific knowledge (29 studies). Research participants went to virtual fieldtrips to distant locations to explore the site and learn the culture (13 studies) or to learn about environmental consequences (7 studies). In virtual fieldtrips to the past, learners observed historical towns and learnt how historical events unfolded (10 studies). In a small number of studies, users became someone else to experience their daily lives as a patient or student (5 studies) or designed 3D objects in virtual environments (5 studies).

Simulation-based IVR applications offered opportunities to practice performance-based tasks in realistic and engaging virtual environments that are safe and cost effective. Medical students practiced surgical procedures or diagnostic protocols (25 studies). Children and adults enacted and learnt the safety protocols in case of fire or earthquake (23 studies). Science students followed step-by-step instructions to become familiar with how to use lab equipment (12 studies) and engineering trainees practiced machine maintenance or operation protocols at virtual job sites (12 studies). Participants in 12 studies practiced presentation or communication skills in front of virtual audiences. Novice and expert drivers drove virtual vehicles in various road conditions (7 studies).

Researchers adopted simple game elements (e.g., scoring system) in IVR applications to better engage learners in learning various topics, such as practicing safety protocols (8 studies), learning science content (7 studies) or maths contents (4 studies), or taking virtual fieldtrips (4 studies).

When the design feature integration was compared against the different types of application types, only minor differences were found with simulations showing slightly more design features compared to virtual world representations (see Table 5).

4.3. Five approaches to learning with IVR

To find patterns in integration of the design features, this review used k-means clustering to identify five distinct clusters amongst

![Table 4](image)

**Correlation between design features.**

<table>
<thead>
<tr>
<th></th>
<th>visual</th>
<th>audio</th>
<th>haptic</th>
<th>interactivity</th>
<th>movement</th>
<th>roles</th>
<th>contexts</th>
<th>challenges</th>
<th>social</th>
</tr>
</thead>
<tbody>
<tr>
<td>visual</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>audio</td>
<td>-0.10</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>haptic</td>
<td>0.25**</td>
<td>0.20**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>interactivity</td>
<td>0.42**</td>
<td>-0.10</td>
<td>0.14*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>movement</td>
<td>0.58**</td>
<td>-0.04</td>
<td>0.40**</td>
<td>0.56**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>roles</td>
<td>0.28**</td>
<td>0.10</td>
<td>0.24**</td>
<td>0.36**</td>
<td>0.33**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>contexts</td>
<td>-0.05</td>
<td>0.19**</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.10</td>
<td>0.50**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>challenges</td>
<td>0.02</td>
<td>0.09</td>
<td>0.04</td>
<td>0.20**</td>
<td>0.14*</td>
<td>0.23**</td>
<td>0.08</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>social</td>
<td>-0.09</td>
<td>0.15*</td>
<td>-0.08</td>
<td>0.16*</td>
<td>-0.03</td>
<td>0.22**</td>
<td>0.17*</td>
<td>0.27**</td>
<td>1</td>
</tr>
</tbody>
</table>

*p < 0.05.

**p < 0.01.

Among pedagogical design features, roles showed a strong correlation with all other design features, except audio. It implies that researchers recognised assigning a specific role as important to help learners feel the ownership of the learning tasks and be engaged in learning. In contrast, two other narrative design features (i.e., contexts and challenges) and the social design feature showed moderate correlations with roles and with one another, but not so much with other technological design features. This observation implies that the pedagogical design features tended to be implemented together, irrespective of the implementation of technological design features.
219 empirical studies. Although all nine design features were used to group the studies, principal component analysis (PCA) was used to visualise the separation of the clusters using three main components (see Fig. 3 and Appendix on PCA results). The studies in each cluster exhibited similar design features when compared against the studies in other clusters. The studies in each cluster also showed similar learning tasks and contexts. Considering that integration of additional design feature in IVR application is costly and time consuming, it can be assumed that researchers would optimise the application development with essential features to achieve their goals. In this sense, the shared design features and learning tasks could represent a shared approach to design IVR-based learning. The five clusters and corresponding approaches were named as follows: watch virtual worlds passively; interact with 3D objects; play a role in context; practice in realistic hybrid environments; and respond to personalised intelligent feedback. The characteristic design features (Table 6) and learning topics (Table 7) of each approach are described below. Hyperlinks to screenshot images are added for the example studies, when available.

4.3.1. Approach #1: watch virtual worlds passively

The studies with the first approach (N = 63) adopted phone-based IVR headsets, in contrast to many IVR studies that adopted high quality graphics equipment to explore new possibilities of the cutting-edge technology in IVR. Because phone-based headsets had limited capacity for motion tracking, the IVR applications lacked interactivity and embodied movement, leading students to passively watch virtual worlds. The main learning tasks tended to be about seeing virtual worlds with audio instructions as an introduction to the topic. The most common learning activity was taking a class of students on a virtual tour with Google Expedition or similar applications to learn about geography or history (e.g., Cheng & Tsai, 2019; Han, 2021) or environmental impacts (e.g., Petersen et al., 2020) or to motivate learning (Yang, Chen, Zheng, & Hwang, 2021). Teachers would give background information before individual students put on the headsets; after virtual exploration, students would engage in class discussions. Another common learning activity was visualisation of abstract science content through watching animations (e.g., Albus, Vogt, & Seufert, 2021). Procedural skill-oriented IVR applications showed step-by-step procedures and asked students to select the right options so that learners become aware of the protocols in medical practices (e.g., Harrington et al., 2018) or science laboratories (e.g., Dunnagan et al., 2020).

For the lack of interactivity and passive learning, the studies with this approach may be critiqued from both technological and pedagogical perspectives. The priority of these studies was the accessibility to reach wider audiences. Phone-based IVR headsets had limited technological features, thus leaving little room for more creative designs, but they had the advantage of easy, low-budget content creation and distribution. Thanks to the accessibility of the phone-based IVR headsets, about a third of the studies in this cluster used IVR as part of class instruction, with primary and secondary school students as well as university students. In some cases, students watched 360° videos and then involved in creative tasks afterwards (e.g., Yang et al., 2021). Utilising the convenience of creating 360° videos, two studies asked students to create 360° videos themselves to practice their foreign language skills (Lin & Wang, 2021) or art museum curation (Nortvig, Petersen, Helsinghof, & Brænder, 2020). Although the IVR applications as product lacked interactivity and other design features, the creation process for students was highly hands-on and reflective. This approach may appeal to educators who wish to experiment with the new technology and integrate it as part of a learning program.

4.3.2. Approach #2: observe 3D objects interactively

The studies with the second approach (N = 43) adopted higher end IVR headsets to offer better 3D visualisation and more interactive learning experiences. In science and engineering, 3D spatial thinking is critical for establishing content knowledge and
skills, and students were given opportunities to interact with 3D objects, albeit what students could do with the virtual objects was limited, such as observing, rotating, and placing it in a place. The most common learning activities was observing 3D anatomical models (e.g., Fairén, Moyés, & Insa, 2020) or organelles in cells (e.g., Parong & Mayer, 2018) to visualise the spatial aspects of science concepts. Other studies involved engineering students or trainee technicians to practice skills in assembling or operating machines (e.g., Winther, Ravindran, Svendsen, & Feuchtner, 2020). The adoption of high-end headsets and expensive computer systems had the advantage of allowing participants to observe 3D objects more interactively, but it also had the limitation of restricted access. The IVR applications were designed to offer supplementary learning experiences for science and engineering students, but the research studies were conducted in university research facilities, without direct connection to formal instructions or interaction with peers. Because physical models or computer simulations were available to achieve the same learning goals, the studies in this cluster compared the effectiveness of IVR against existing learning materials in knowledge acquisition or retention, making them excellent candidates for comparative meta-analysis. However, due to limited implementation of pedagogical design features, the learning in IVR tended to be dry and individualistic without contextualisation, personalised intellectual challenges, or social interactions.

### 4.3.3. Approach #3: play a role in context

The studies with the third approach (N = 63) gave specific roles to learners so that they became someone else to experience certain things. For example, students became historical actors themselves, a soldier experiencing the battles of World War 2 (Calvert & Abadia, 2020) or Lenin leading the Russian Revolution (Slater et al., 2018). Such virtual fieldtrip applications were more interactive and elaborate compared to the studies with the ‘watch virtual world’ approach. Another common study area with this approach was safety training for adults and children. With IVR headsets on, participants became someone who were caught in an earthquake aftermath and learnt how to evacuate a building (e.g., Feng, González, Mutch, et al., 2020). In these applications, the tasks often demanded the recognition or evaluation of the given situation when practicing protocols. For science content learning, engaging storylines and game

---

### Table 6

Integration of design features of five approaches in educational IVR studies.

<table>
<thead>
<tr>
<th>Design approaches</th>
<th>N</th>
<th>visual</th>
<th>audio</th>
<th>haptic</th>
<th>interactivity</th>
<th>movement</th>
<th>roles</th>
<th>contexts</th>
<th>challenges</th>
<th>social</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watch virtual worlds passively</td>
<td>63</td>
<td>1.19 (40)</td>
<td>1.97 (31)</td>
<td>1.00 (90)</td>
<td>1.38 (49)</td>
<td>1.05 (21)</td>
<td>1.30 (46)</td>
<td>1.65 (48)</td>
<td>1.33 (51)</td>
<td>1.37 (49)</td>
</tr>
<tr>
<td>Observe 3D objects interactively</td>
<td>43</td>
<td>2.88 (32)</td>
<td>1.65 (53)</td>
<td>1.07 (26)</td>
<td>1.88 (32)</td>
<td>1.74 (49)</td>
<td>1.16 (37)</td>
<td>1.07 (26)</td>
<td>1.05 (21)</td>
<td>1.00 (00)</td>
</tr>
<tr>
<td>Play a role in context</td>
<td>63</td>
<td>2.75 (44)</td>
<td>1.92 (52)</td>
<td>1.14 (35)</td>
<td>1.86 (35)</td>
<td>1.65 (60)</td>
<td>1.98 (13)</td>
<td>2.10 (35)</td>
<td>1.30 (46)</td>
<td>1.29 (46)</td>
</tr>
<tr>
<td>Practice in realistic hybrid environments</td>
<td>17</td>
<td>2.88 (33)</td>
<td>2.06 (24)</td>
<td>2.76 (44)</td>
<td>2.06 (24)</td>
<td>2.88 (49)</td>
<td>2.00 (00)</td>
<td>1.59 (31)</td>
<td>1.71 (69)</td>
<td>1.18 (39)</td>
</tr>
<tr>
<td>Respond to personalised intelligent feedback</td>
<td>33</td>
<td>2.73 (57)</td>
<td>1.94 (50)</td>
<td>1.03 (17)</td>
<td>2.27 (45)</td>
<td>2.27 (45)</td>
<td>1.85 (36)</td>
<td>1.58 (36)</td>
<td>2.21 (65)</td>
<td>2.00 (56)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>219</td>
<td>2.33 (84)</td>
<td>1.89 (46)</td>
<td>1.20 (53)</td>
<td>1.80 (50)</td>
<td>1.68 (70)</td>
<td>1.61 (49)</td>
<td>1.65 (55)</td>
<td>1.42 (60)</td>
<td>1.35 (52)</td>
</tr>
</tbody>
</table>

---

### Table 7

Common learning tasks observed in five approaches in educational IVR studies.

<table>
<thead>
<tr>
<th>Design approaches (# of studies)</th>
<th>Description of common learning tasks and implemented design features</th>
<th>Common learning topics* (# of studies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watch virtual worlds passively (63)</td>
<td>Learners watched 360° videos or follow step-by-step instructions. With lower-end IVR headsets, interactivity and movement were limited.</td>
<td>virtual tour (17), science visualisation (10), medical procedures (8), and science lab procedures (7)</td>
</tr>
<tr>
<td>Observe 3D objects interactively (43)</td>
<td>Learners observed and interacted with virtual objects following explicit instructions. Visual, interactivity, and movement were better implemented, but the integration of pedagogical features was limited.</td>
<td>science visualisation (18), and engineering procedures (7)</td>
</tr>
<tr>
<td>Play a role in context (63)</td>
<td>Learners assumed a role to experience events or practice a specific performance-oriented task in sequence. Strong contextualisation was observed.</td>
<td>safety training (18), virtual fieldtrips (13), medical procedures (9), science visualisation (5), science lab procedures (4), and empathy (4)</td>
</tr>
<tr>
<td>Practice in realistic hybrid environments (17)</td>
<td>Learners practiced their skills in a simulated hybrid physical-virtual environment with realistic haptic force feedback. All technological features were well implemented with some pedagogical features.</td>
<td>surgical procedures (7), and driving (5)</td>
</tr>
<tr>
<td>Respond to personalised intelligent feedback (33)</td>
<td>Learners completed a series of tasks often through multiple attempts with personalised intelligent feedback or mediated social interactions. Strong pedagogical features were observed.</td>
<td>communication (8), science visualisation (5), design (4), and physical education (4)</td>
</tr>
</tbody>
</table>

* Excluded learning topics with three or less studies.
elements were integrated, for example, students becoming an immune system soldier to fight off bacteria with varying weapons and intensity (Zhang et al., 2019). Nursing students became a geriatric patient with macular degeneration and hearing loss (Buchman & Henderson, 2019) to build empathy and communication skills, and pre-service teachers became a student drug user (Stavroulia et al., 2019) to experience their stories in virtual environments.

The participants in these studies did not have full-body avatars. Yet, seeing virtual worlds as someone and navigating the virtual worlds with specific goals compelled them to complete the tasks with more emotional and intellectual engagement. IVR researchers recognise that experiencing virtual worlds as someone else is one of the most powerful features of IVR (Slater, 2017; Yee & Bailenson, 2007). It is worth pointing out that learners did not have much freedom to change the course of actions or events within the virtual environments, unfortunately. They tended to follow scripted pathways. Some researchers experimented with the impact of learners’ autonomous actions in IVR through self-paced vs. scripted virtual tours (Checa & Bustillo, 2020). Educational researchers may want to explore various ways to integrate more interactivity and other design features to this approach for more active learning experiences.

4.3.4. Approach #4: practice in realistic hybrid environments

Among three sensory design features, haptics was the least adopted in educational IVR applications as noted above (see Fig. 2). The studies with the fourth approach (N = 17) strengthened the technological features by setting up hybrid haptic environments and integrating relevant embodied movement to offer a highly immersive, realistic experience for learners. Two learning areas stood out in this approach: surgical procedure and driving practice. They adapted existing hybrid simulators for 2D computer environments to create realistic hybrid environments in IVR. With laparoscopic grippers and resistance simulators, surgical training applications offered realistic tactile feedback (e.g., Barré et al., 2019; Frederiksen et al., 2020). With a driver seat with wheel and brake on a platform, driving simulations responded to users’ input to offer realistic training opportunities (e.g., Ropelato et al., 2018). In these studies, body movement or muscle memory was the critical aspect of learning.

With customised extra equipment to enhance realism, these studies set up the most technologically advanced virtual environments, testing the technical boundaries of IVR-based learning. Such experimentation unfortunately did not extend to implementing other pedagogical design features. Given that such resource-rich virtual learning would be difficult to replicate in other learning areas especially in formal education settings, their focus on technical aspects may be fitting for the purpose, at least for the time being.

4.3.5. Approach # 5: respond to personalised intelligent feedback

The last batch of studies (N = 33) did not integrate the technological features as well as the studies above. Their defining features were what was missing from the other studies. In the studies for communication or presentation skill development, for example, participants were given opportunities to practice their skills over multiple sessions in front of a virtual audience (e.g., Gruber & Kaplan-Rakowski, 2020). Depending on the learners’ performance, the virtual audience’s facial expressions, postures, or verbal responses were adjusted (e.g., Herrero & Lorenzo, 2020; Lugrin et al., 2016; Real et al., 2017). After their virtual performance, participants also received automatically generated speech analysis charts for personalised feedback (e.g., van Ginkel et al., 2019). Because learners perceived that they were presenting in front of intelligent virtual humans, the experience felt realistic but safe to practice their skills and improve on their performance over time.

The studies for visualisation of abstract concepts, in contrast, made peer-to-peer interactions and collaborative meaning making as their distinctive design feature. Two or more students cohabited virtual environments, physically standing next to each other (e.g., Won, Mocerino, Tang, Treagust, & Tasker, 2019) or apart (e.g., Sašinka et al., 2019), exploring virtual environments, discussing their ideas, and completing the learning tasks collaboratively. Arts and design education studies took yet another direction. They offered neither multiple opportunities nor social interactions. Instead, they offered creative opportunities: instead of following a script, students designed new 3D artefacts in virtual environments (Ho, Sun, & Tsai, 2019; Yang et al., 2018) or design and 3D-print the artefacts (Guan, Wang, Chen, Jin, & Hwang, 2021).

The studies with this approach tended to have open-ended learning tasks, interacting with others productively. Collaborative learning model was integrated by overcoming multiple technological challenges, but not necessarily with sophisticated technological solutions. Because intelligent tutoring system was not available to encourage constructive learning, researchers functioned as puppets behind the scenes (e.g., Herrero & Lorenzo, 2020; Lugrin et al., 2016; Real et al., 2017). As real-time networking and synchronisation of multiple learners was difficult, researchers limited the number of participants, simplified the interface, and encouraged discussion (e.g., Won et al., 2019; Sašinka et al., 2019).

5. Discussion

5.1. The landscape of IVR-based learning design

With the emergence of easy-to-use video conversion tools and IVR application authoring tools (e.g., Unity™ and Unreal™), it is becoming easier to create immersive virtual learning environments. Improved and cheaper IVR hardware also increases the feasibility to implement it for teaching and learning. Without huge research grants, educational researchers—many in collaboration with professional game developers—are creating and experimenting with IVR. At this exciting time of evolving technology, it is crucial to examine how researchers are designing IVR applications to support learning because the educational values of a new technology are not just based on the technological capabilities themselves but on how it is designed and implemented to encourage learning (Wu et al., 2013).

To understand how educational IVR applications were designed, this review analysed contemporary empirical studies in relation to
their integration of technological and pedagogical design features. For the individual design features, many researchers utilised the technological advantages of IVR to establish realistic, immersive 3D virtual environments by adopting high end IVR headsets and integrating interactive interface and embodied movement. In contrast, the pedagogical design considerations were observed less frequently in educational IVR studies, such as contextualisation of the tasks, collaborative peer-to-peer interactions, and multiple learning opportunities. This observation may be aligned with the common critique of IVR studies for focusing on the development and usability testing of IVR applications rather than the evaluation of learning outcomes (e.g., Hamilton et al., 2021; Radianti et al., 2020).

The limited integration and evaluation of pedagogical features, however, may exist not simply because researchers were not considering pedagogical approaches, but because they were still grappling with evolving technological capabilities (Coban, Bolat, & Goksu, 2022; Garzón, Baldiris, et al., 2020). For example, due to the weight of headset and the potential side effects of dizziness or headache, researchers tended to limit IVR usage to a short period of time, which made it difficult to engage learners in a comprehensive inquiry-based learning or a quest-like learning task. To encourage collaborative social constructivist learning, as an alternative pedagogical approach, multi-player design needs to be effectively integrated, but it remains challenging in many IVR platforms. Such technical challenges may lead educational researchers to focus on different aspects of IVR-based learning than traditionally valued pedagogical considerations or approaches.

To understand how researchers are navigating the emerging educational capabilities (and challenges) of IVR, this review used cluster analysis (a machine learning protocol) and identified five common approaches. Each design approach appeared to indicate different priorities of the researchers when designing IVR-based learning. Some researchers prioritised the accessibility of the learning material over all else and adopted cheaper IVR headsets so that all students in class can experience virtual worlds. This choice unfortunately limited the interactivity and the scope for integrating other design features. Some other researchers, in contrast, focused on providing a highly immersive experience that IVR could offer and organised the setup to be realistic, interactive virtual environments. This choice for additional hardware and software features greatly reduced the learners’ access to the IVR learning as results. Yet other researchers valued the emotional engagement or social interactions above other considerations. They took a middle road for accessibility and realistic interactions and instead integrated other design features, such as intriguing contexts or peer interactions.

Because of such diverse priorities, even the ‘same’ learning topics were approached differently in IVR environments, and learners performed dissimilar learning tasks. For example, for the topic of learning medical procedures, the studies with the first approach (watch virtual worlds passively) would show a video of a surgical operating theatre (Arents, Pieter, Struben, & van Stralen, 2021) so that undergraduate student cohorts became familiar with the procedures, but no ‘hands-on’ aspect was integrated in the activity. The studies with the third approach (play a role in context) would ask learners to take the role of health care workers and perform a series of hands-on tasks such as positioning patients and taking X-rays (e.g., O’Connor et al., 2021) so that university students could envision themselves doing the job, but without feeling realistic tactile feedback from the virtual patients. In the studies with the fourth approach (practice in realistic hybrid environments), junior doctors simulated operating on virtual patients handling realistic surgical equipment in a virtual operating theatre (e.g., Barré et al., 2019; Frederiksen et al., 2020).

The topic of visualisation of abstract science concepts could serve as another example to illustrate the diversity in approaching the same learning topic. The studies with the first approach (watch virtual worlds passively) would show an animation of cell organelles with a voice-over (Huang, 2022) while the studies with the second approach (observe 3D objects interactively) would let learners hold cell organelles and observe them from multiple angles (Parong & Mayer, 2018). Students in the studies with the third approach (play a role in context) would become an immune system soldier fighting off bacteria in human body to learn how human immune system works (Zhang et al., 2019) while students in the studies with the fifth approach (respond to personalised intelligent feedback) would explore the inside of a protein enzyme with their peers and construct knowledge on catalytic reaction collaboratively (e.g., Won, Mocerino, Tang, Treagust, & Tasker, 2019).

5.2. Implications and suggestions for future studies

Educational researchers time and again emphasised the importance of establishing design principles and integrating pedagogy when designing and evaluating technology-supported learning activities (Barab et al., 2010; Klopfer, 2008; Mikropoulos & Natsis, 2011). By including both technological and pedagogical considerations as part of the design features of IVR-based learning activities, this review highlighted educational and practical aspects that education designers and researchers have considered so far and what they could consider further in future. Of course, the nine design features are not an exhaustive list and thus the five approaches are not set in stone—they may be expanded and refined as more ingenious usage of the IVR technology is introduced and tested for education. Yet, the design features from this review are tangible and concrete for educational IVR researchers to refer to when they plan learning goals and implement corresponding design features in IVR learning activities. The design approaches scoped in this review are also sufficiently diverse and comprehensive for educational researchers to base their IVR-based learning activities on as they negotiate competing design decisions around the constraints of the learning environments, the resources, and the technology itself.

The identification of five common approaches in empirical studies on IVR also offers alternative directions in systematic literature reviews. Firstly, the review suggests that in the analysis of empirical studies on IVR, more attention is needed in the way the learning tasks were designed. Existing systematic literature reviews and meta-analyses have tended to categorise empirical studies based on learner demographics, learning objectives or topics and to compare the efficacy of IVR on educational outcomes (Coban et al., 2022; Hamilton et al., 2021; Jensen & Konradsen, 2018; Radianti et al., 2020; Wu, Yu, & Gu, 2020). However, this review reveals that even with the same learner demographics and learning topics, empirical studies could offer vastly different IVR-based learning experiences. Placing such diverse learning experiences into one group and comparing their learning outcomes against other studies on different topics may not be as productive as educators would hope. Instead, in the future, researchers may wish to consider different approaches
to design and learning in their analysis of educational IVR studies.

Secondly, this review demonstrates the benefits of an open, exploratory method for reviewing empirical studies with new educational technology, such as IVR. In previous systematic literature reviews, different pedagogical approaches were commonly drawn a priori from theoretical literature (e.g., collaborative learning, situated learning, inquiry-based learning, and multimedia learning), and applied to analyse empirical studies (e.g., Garzón, Baldiris, et al., 2020; Radianti et al., 2020). IVR-based learning environments may progress and allow educators to achieve more social constructivist educational goals, such as helping learners understand the complex interactions of real life scenarios and solve real life problems through collaboration and communication (Abdul Jabbar & Felicia, 2015; Chen et al., 2018; Garzón, Baldiris, et al., 2020; Sung, Yang, & Lee, 2017). This review recognised that current IVR technology might not yet be well aligned with such pedagogical approaches. This review inductively identified five common approaches to learning through a clustering technique and examination of the implemented learning task designs in IVR. These five approaches illustrated how researchers were exploring and utilising the emerging technology to realise their educational goals while negotiating various constraints related to resources, environments, and technology. Such illustrations better reflect fast-growing research areas, such as educational IVR, and have the potential to support further research on educational possibilities of new technology.

Lastly, thanks to the extensive coverage of the empirical studies, this review portrayed a distinctively different landscape of educational IVR than those by other systematic reviews or meta-analyses. Systematic reviews or meta-analyses typically include a small number of educational studies. For reviews on educational IVR (Coban et al., 2022; Di Natale et al., 2020; Hamilton et al., 2021; Jensen & Konradsen, 2018; Radianti et al., 2020), the number of reviewed empirical studies ranged from 18 to 39 studies with the median of 27.5 studies. Because of the small number, less than a handful of empirical studies represented a whole demographic group (e.g., k-12 students) or a certain discipline area (e.g., architecture) and they were applauded as the most effective educational use of IVR (Coban et al., 2022). In addition, existing literature reviews tended to include a certain type of studies with comparative experimental design, comparing learning the ‘same’ content across various platforms such as IVR, computer display, physical models, and lectures (Coban et al., 2022). In contrast, this review analysed a much larger sample of empirical studies (219 studies) to include diverse discipline areas, learning tasks, and research designs while adopting a quality measure with the citation numbers. Because of the comprehensive coverage, this review was able to identify the emerging patterns in design and learning approaches in contemporary IVR studies beyond what is currently available in the extant literature. In future systematic reviews, researchers may want to consider adopting similar data collection and analysis methods, such as the web scraping techniques for identifying and screening a large volume of literature, and the clustering analysis for surveying the trends in empirical studies.

5.3. Limitations

Although the authors of this systematic literature review made a concerted effort to include many impactful research studies, we screened out a significant number of studies using the number of citations due to the sheer volume of recent IVR literature. Some researchers use a broader term such as ‘Virtual Reality’ rather than ‘immersive Virtual Reality’ when referring to similar technology. Some others may use a different phrase to refer to the equipment rather than ‘HMD’ or ‘headset’. Consequently, there may be some important studies not included. However, the range of studies included in this manuscript encompasses diverse research across disciplinary areas and learning contexts to represent the status of IVR-based education studies. In addition, by reviewing the studies with the potential to impact future studies, this review may be illustrative in forecasting future research directions. For more detailed literature reviews within a discipline- or learning context-specific realm (e.g., Matovu et al., 2022), researchers may wish to use the similar data collection and analysis methods combining broader search keywords, such as Virtual Reality and presence, with context-specific keywords, such as university science laboratory procedures or primary school social skill development, to identify common approaches and learning tasks in educational IVR studies for the specific contexts.

The clustering method effectively captured common approaches within each cluster and identified some important differences between clusters, as illustrated above. However, the machine learning technique unfortunately did not always succeed in grouping studies with similar approaches. Because the visual design feature plays a critical role in distinguishing the first cluster from the rest, the same IVR application of watching virtual worlds passively could be grouped as watching virtual worlds with lower-end headsets or as playing a role in context with higher-end headsets in some cases (e.g., Rupp et al., 2019). Despite these limitations, the clustering technique captured important patterns in educational IVR application designs.

The authors of this review originally attempted to document how the implemented design features impact the learning outcomes. Unfortunately, this review did not investigate the links between implemented design features and learning outcomes or between design approaches and learning outcomes. It is partly because many of 219 studies reported learners’ positive perception of IVR experience rather than systematically evaluating the learning outcomes as noted in other systematic reviews (Hamilton et al., 2021; Radianti et al., 2020). Instead, this review focused exclusively on documenting the diversity in the approaches to learning in educational IVR studies. Future investigators may wish to establish the link between the implemented design features, learning approaches, and the learning outcomes.
When implementing new technology, educators consider many other aspects than what we discussed in this review, such as: How much background is assumed for learners (both content and the technology); how students can be supported before, during, and after the activity to make learning productive; and what assistance do I get as the implementer of the technology. Future studies may look at different forms and levels of support provided to overcome the novelty effect or design flaws and make IVR-based learning meaningful and relevant.

6. Conclusion

A growing number of educational researchers are investigating the feasibility and the potentials of IVR for educational use. Although numerous empirical studies report students’ positive perceptions towards IVR-based learning, it is unclear what makes IVR-based learning effective and how educational researchers are utilising the unique features of IVR. This systematic review addresses this gap in the literature by examining how educational researchers have designed IVR learning environments in contemporary educational IVR studies. To document various designs of educational IVR applications, this review adapted and expanded Dede’s (2009; 2017) immersive interface design to have both technological and pedagogical design features. From the analysis of the levels of integration of design features in 219 empirical studies published between 2016 and 2021, this review observed that the technological features (e.g., high-quality graphics and interactive interface) were more readily implemented to enhance physical presence. In contrast, pedagogical features (e.g., contextualisation and social interactions) for learner engagement were less frequently observed. Further clustering analysis revealed that such low levels of integration for pedagogical features did not necessarily mean that researchers focused more on the technological aspects than on the pedagogical aspects. Rather, this analysis indicated that, because it is costly and time consuming to integrate all design features, researchers integrated different combinations of design features depending on their educational objectives, priorities, and environmental constraints. This review identified five common approaches to designing IVR-based learning tasks. Some researchers chose lower-end IVR hardware to involve many students at the same time and included additional class activities to situate and complement IVR tasks. Other researchers had limited connections to formal education and instead aimed at offering intense emotional involvement as a stand-alone activity through integrating higher-end IVR hardware and interesting storylines or personalised learning opportunities. Different from other systematic literature reviews on educational IVR studies, this review clearly illustrated the diversity in the approaches to designing IVR-based learning. We urge educational researchers to consider the existing approaches and re-envision how they design IVR-based learning tasks to achieve their pedagogical objectives and how they evaluate the educational possibilities of IVR.

Credit author statement

Mihye Won: conceptualisation, methodology, formal analysis, writing – original draft, Dewi Ayu Kencana Ungu: formal analysis, Henry Matovu: formal analysis, David F. Treagust: writing – review & editing, Chin-Chung Tsai: conceptualisation, writing-review, Jungho Park: formal analysis, visualisation, Mauro Mocerino: writing-editing, Roy Tasker: writing-editing.

Data availability

The data is available as online supplementary material.

Acknowledgements

Funding is provided by the Australian Research Council Discovery Project, Using Immersive Virtual Reality to Enhance Students’ Science Visualisation (DP190100160).

Appendix A. Principal Component Analysis

Principal component analysis (PCA) was conducted to understand how different design features were adopted in relation to one another and what features were the most distinguishing aspects of IVR applications. PCA is an effective and widely used exploratory data analysis method (Jolliffe, 2005). PCA reduces the dimensionality of the data by creating an alternative set of explanatory variables called principal components and rotating the data (Varimax) to make the principal components as the new axes. From the scree plot of PCA (Fig. A1), four principal components were chosen for further analysis because (a) four principal components accounted for 75% of explained variance of the data; (b) four principal components were larger than the average explained variance; and (c) the remainder of principal components offered relatively little to explain the variance (Jolliffe, 2002). The loading of each variable to the four selected principal components was evaluated subsequently (Table A1). In the main manuscript, all 219 studies were plotted on the axes of the three principal components, PC1, PC2 and PC3 (see Fig. 3).
Principal Component (PC) 1 had strong positive contributions from three technological design features—visual, embodied movement, and interactivity. This means, PC1 separates IVR applications with relatively high-level visual and actional design features from the ones with relatively low-level integrations of those technological design features. The studies with high values on the PC1 axis (see Fig. 3) include the studies with a strong emphasis on embodied movement (Clusters #4 & #5), such as driving simulations (e.g., Lang, Wei, Xu, Zhao, & Yu, 2018; Ropelato et al., 2018), golf putting simulations (e.g., Harris et al., 2020), and surgery simulations (e.g., Barré et al., 2019), and also art design applications (e.g., Guan et al., 2021; Ho et al., 2019; Yang et al., 2018). These studies adopted high-end IVR headsets with robust user interface to allow intuitive interactions and natural body movement. In contrast, the studies with low values on the first axis include the studies showing 360° movies (Cluster #1) on medical procedures (e.g., Chang et al., 2019; Wang et al., 2020; Sattar et al., 2020) or for virtual tours (e.g., Griffin & Muldoon, 2020; Nelson, Angraini, & Schlüter, 2020). Those studies focused on delivering new information through 360° videos without much interactivity or embodied movement in phone-based IVR headsets.

Challenge, context, social interactions, and roles—all four pedagogical design features—contributed strongly to PC2. The studies with high values on the PC2 axis offered multiple learning opportunities in IVR for learners to engage and improve on their skills in specific contexts, such as communication skills in classrooms, doctor’s offices, or streets. These studies supported learning through direct or indirect social interactions (e.g., Barreda-Angeles, Aleix-Guillaume, & Pereda-Baños, 2020; Herrero & Lorenzo, 2020; Liaw, 2019; Lugrin et al., 2016; McCleery et al., 2020; McFaul & FitzGerald, 2019; Real et al., 2017). The studies with low values on the same axis include the studies for 3D visualisation of science topics, such as human anatomy (e.g., Chen et al., 2020; Ekstrand et al., 2018; Napa, Moore, & Bardyn, 2019; Parkhomenko et al., 2019; Pfeiffer et al., 2018; Weyhe, Uslar, Weyhe, Kaluschke, & Zachmann, 2018), where learners manipulate 3D objects interactively (Cluster #2). Those studies tended to adopt high-end IVR headsets, but did not offer specific roles, contexts, or missions to accomplish. There was no scope for social support or progressive challenges in learning.

Both technological and pedagogical design features contributed to PC3, some positively and some negatively. The loading distribution to PC3 may appear confusing, but it differentiates the studies with a strong emphasis on empathy but with limited embodied movement (Cluster #3). The studies with high values on the PC3 axis include the studies aiming to build empathy (e.g., Calvert & Abadia, 2020; Herrera et al., 2018; Slater et al., 2018; Stavroulia et al., 2019). These studies set up concrete contexts where learners experience an event as historical or social actors themselves. Although these studies used high-end IVR headsets, embodied movement was limited. The studies with low values on the third axis include the studies focusing on embodied movement but without strong storylines or contextual information (Clusters #4 & #5), such as driving simulations (e.g., Ebnali, Lamb, Fathi, & Hulme, 2021; Lang et al., 2018; Ropelato et al., 2018) or art design applications (e.g., Guan et al., 2021; Ho et al., 2019; Yang et al., 2018). These studies often gave multiple opportunities to complete the given tasks.

Two remaining technological design features contributed to PC4, haptic and audio. The studies with high values on this axis were the studies with hybrid environments in Cluster #4, such as surgery simulators with grippers (e.g., Frederiksen et al., 2020; Huber et al., 2019; McFaul & FitzGerald, 2019; McCleery et al., 2020).

<table>
<thead>
<tr>
<th>Principal component analysis (loadings to four principal components)</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>visual</td>
<td>0.69</td>
<td>-0.31</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>movement</td>
<td>0.56</td>
<td>-0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>interactivity</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>challenges</td>
<td>0.53</td>
<td>-0.44</td>
<td>-0.35</td>
<td></td>
</tr>
<tr>
<td>contexts</td>
<td>0.47</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>social</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>roles</td>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>haptic</td>
<td></td>
<td></td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>audio</td>
<td></td>
<td></td>
<td></td>
<td>0.38</td>
</tr>
</tbody>
</table>

* the list of design features is rearranged based on the loading amount.
et al., 2018) or crane operation simulators with switches and handles (e.g., Song, Kim, Kim, Ahn, & Kang, 2021). These studies had a hybrid physical-virtual environment setup to offer realistic haptic feedback. The studies with low values on this axis include the studies with low haptic or audio design features, such as art design applications in Cluster #5 (e.g., Guan et al., 2021; Ho et al., 2019; Kim et al., 2020; Yang et al., 2018) or safety training in Cluster #3 (e.g., Dixon, Miyake, Nohelty, Novack, & Granpeesheh, 2019; Çakıroğlu & Gokoğlu, 2019). These applications allowed multiple opportunities to craft their skills.

Appendix B. Supplementary materials

Supplementary materials to this article can be found on https://doi.org/10.1016/j.compedu.2022.104701.

References


Ochs, M., Mestre, D., de Montcheuil, G., Pergandi, J.-M., Mulrow, C. D., et al. (2019). Training doctors with low haptic or audio design features, such as art design applications in Cluster #5 (e.g., Guan et al., 2021; Ho et al., 2019; Kim et al., 2020; Yang et al., 2018) or safety training in Cluster #3 (e.g., Dixon, Miyake, Nohelty, Novack, & Granpeesheh, 2019; Çakıroğlu & Gokoğlu, 2019). These applications allowed multiple opportunities to craft their skills.


Appendix C. List of 219 Studies Reviewed in the Study


*Taçgın, Z. (2020). The perceived effectiveness regarding Immersive Virtual Reality learning environments changes by the prior knowledge of learners.


*Sim, L., Goh, R. K., & Tse, P. (2018). Designing a virtual learning environment for automating the learning of primary students using computational thinking.


