frontiers in Education

TYPE Perspective
PUBLISHED 28 August 2026
DOI 10.3389/feduc.2026.1868028

OPEN ACCESS

EDITED BY
Koen Smit,
HU University of Applied Sciences Utrecht, Netherlands

REVIEWED BY
Sandeep Chouhan,
University of Birmingham, United Kingdom
Francesca Finestrone,
University of Foggia, Italy

*CORRESPONDENCE
Pedro Acevedo
acevedo@uncw.edu

RECEIVED 28 April 2026
REVISED 04 August 2026
ACCEPTED 18 August 2026
PUBLISHED 28 August 2026

CITATION
Acevedo P, Magana AJ, Benes B and Mousas C (2026) Reinterpreting multimedia learning principles for extended reality: a perspective on XR learning design.
Front. Educ. 11:1868028.
doi: 10.3389/feduc.2026.1868028

Reinterpreting multimedia learning principles for extended reality: a perspective on XR learning design

Pedro Acevedo[1]*, Alejandra J. Magana[2], Bedrich Benes[3] and Christos Mousas[2]

[1]Department of Computer Science, University of North Carolina Wilmington, Wilmington, NC, United States, [2]School of Applied and Creative Computing, Purdue University, West Lafayette, IN, United States, [3]Department of Computer Science, Purdue University, West Lafayette, IN, United States

Extended reality (XR) enables users to interact with and control simulated 3D elements, creating opportunities to support meaningful learning experiences, but it also introduces cognitive load due to its novel interactions. Researchers and designers have attempted to address these challenges by manipulating the user experience through the design of the virtual environment or by varying the level of interaction complexity. While traditional media benefit from established instructional frameworks, immersive media such as XR lack similarly refined design guidelines. In this perspective article, we argue that Mayer's multimedia learning principles cannot be directly transferred to XR learning environments without reinterpretation, as XR introduces distinctive cognitive and interaction demands, including embodiment, spatial interaction, presence, agency, and environmental complexity. From this perspective, we organize XR design considerations around cognitive load processes and discuss how these choices may either support or hinder learning. We then propose key considerations for designing effective XR-based learning experiences, including content representation (e.g., 2D and 3D elements, color, positioning), virtual environment settings (e.g., background elements, haptic feedback, sound, voice interactions), and the structure of learning materials (e.g., pre-training materials, interactions and tasks, assessment).

KEYWORDS
augmented reality, design, extended reality, instructional design, multimedia learning principles, virtual reality

1 Introduction

Extended reality (XR) as a medium for interactive and engaging experiences has been used for entertainment, learning, and training. In education, XR can support learning gains and foster engagement when experiences are carefully structured and intentionally designed. Unlike traditional computer-based programs that rely on 2D elements and limited embodied interaction, immersive media introduce distinct affordances through head-mounted displays (HMDs) or see-through displays, including presence, immersion, and embodiment. Effectively leveraging these affordances for educational purposes poses a significant challenge for instructional designers, developers, and researchers.

Researchers and designers have tried to tackle that challenge by adjusting variables under their control, such as the design of the virtual environment, content complexity, and interaction mechanisms. While traditional media have long benefited from established frameworks, such as multimedia learning principles (Mayer, 2009), to reduce cognitive overload, immersive media, such as XR, lack similarly refined guidelines. We address this need by examining XR-specific design elements through the lens of multimedia learning principles to inform effective learning experiences. Our central position is that XR learning design should not simply apply multimedia learning principles developed for screen-based media. Instead, these principles need to be reinterpreted in relation to XR-specific affordances and constraints, particularly embodiment, spatial interaction, environmental complexity, and multimodal feedback.

1.1 Why XR?

The growing adoption of immersive media reflects their transition from specialized applications toward more widespread everyday use. Currently, relatively low HMD costs (e.g., Meta Quest 3S, released in October 2024, costs $299), ease of development, and increasingly natural interaction techniques (e.g., improved hand-tracking and passthrough) have contributed to the broader adoption of immersive technologies across diverse user groups. This also includes low-end devices such as cardboard VR (e.g., Google Cardboard), which enable smartphones to provide stereoscopic views that simulate immersive passive visualizations. This adoption extends to training and education, where XR is leveraged to enhance learning opportunities. Through XR, designers can create experiences that make invisible or complex phenomena perceptible, simulate unsafe or otherwise inaccessible environments, and support hands-on activities within controlled settings (Kaplan-Rakowski et al., 2023). Advances in artificial intelligence (AI) further enable more natural and engaging XR interactions. Current trends include intelligent, hands-free AI assistants that help users navigate their surroundings, as well as AI-powered XR scenarios in which users can have informed and in-depth conversations with virtual agents (e.g., pedagogical agents).

In this regard, can XR be more effective than traditional media solutions? The answer depends largely on how it is designed and implemented. This medium offers multiple advantages, including the ability to simulate real-world scenarios with natural interactions. Thus, learners can perform in ways similar to those in real-life settings, since XR enables embodied interactions (e.g., grasping, pointing, or spatial reasoning) that mirror real-world physical experiences. With XR, designers can create compelling and meaningful learning experiences that integrate multimodal elements to complement their content. However, there is a trade-off between the amount of visual, auditory, or other sensory feedback required and the extent to which the learners' overall surroundings are altered. This can disrupt learning by overwhelming learners' cognitive processes, causing an overload that hinders the learning objective and leading users to focus on the interaction rather than the concepts.

1.2 Cognitive processes during learning

According to Mayer (2009), working memory capacity is limited, requiring learners to manage competing cognitive demands during learning. The demands of learning include extraneous, essential, and generative processing. Extraneous processing refers to the processing of content not related to the instructional goal, including elements that are not part of the learning content. This cognitive process goes hand in hand with the quality of instructional design. Essential processing is responsible for mentally translating the learning material into working memory, which is related to the complexity of the concept and its pace. Finally, generative processing refers to the cognitive process that aims to make sense of and link the incoming information. This process also depends on the learner's motivation to understand the activity.

As these processes rely on the same limited working memory, they compete with one another. Consequently, managing extraneous, intrinsic, and germane processing during learning is essential to prevent cognitive overload. This perspective aligns with *cognitive load theory* (Makransky and Petersen, 2021), which distinguishes among intrinsic load (i.e., the complexity of new information), extraneous load (i.e., processing caused by irrelevant or distracting elements), and germane load (i.e., the integration of new information with prior knowledge).

Working memory management was one of the motivations for Mayer's development of the multimedia learning theory. His framework targets how people learn better when words and pictures are used together than words alone, resulting in meaningful learning. These proposed principles aim to support learners' memory through dual channels (i.e., visual and auditory) and across sensory memory, working memory, and long-term memory. Within this framework, learners select, organize, and integrate information from the learning materials. When considering immersive experiences, multimedia learning principles appear applicable in certain contexts due to the visual and auditory elements they encompass (see Figure 1). However, their relevance is limited, as XR also introduces factors such as agency and presence that extend beyond traditional multimedia frameworks. Prior studies have examined the use of these principles in XR learning, yielding mixed findings regarding their applicability (Çeken and Taşkın, 2022). Since these principles do not account for XR-specific dimensions that may be supported to varying degrees across XR systems, such as embodiment (i.e., how learners experience and use their virtual body and actions), presence (i.e., the sense of "being there" in the virtual environment), and agency (i.e., the sense of control over actions and their effects), there is a clear need to develop and adapt design guidelines tailored to immersive learning environments.

1.3 The need to reinterpret multimedia learning principles for XR

Several frameworks and reviews have attempted to explain how XR supports learning. For example, the Cognitive Affective Model of Immersive Learning (CAMIL) proposed by Makransky

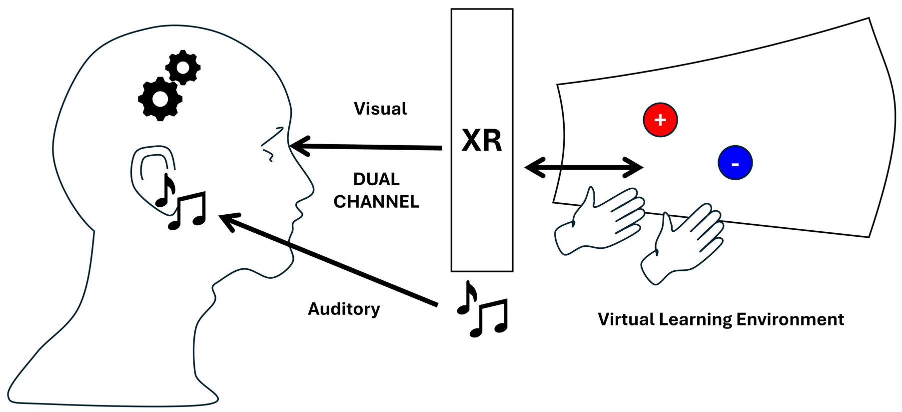

FIGURE 1
Multimedia learning principles within XR environments, illustrating the interaction between visual and auditory channels and immersive virtual content.

and Petersen integrates technological affordances such as immersion, agency, and representational fidelity with factors including cognitive load, motivation, embodiment, and self-regulation (Makransky and Petersen, 2021). Other frameworks have explored XR learning through constructivist, situated, embodied, and student-centered perspectives, emphasizing authentic interaction, collaboration, engagement, and learner autonomy within immersive environments (Wee et al., 2025).

Prior reviews have also highlighted several pedagogical benefits of XR, including increased motivation, engagement, spatial understanding, experiential learning, and improved retention across domains such as science, medicine, engineering, and language learning (Wee et al., 2025). However, these frameworks focus primarily on theoretical relationships, domain-specific applications, or broad pedagogical outcomes rather than explicitly addressing design considerations for XR learning experiences from an instructional designer's perspective.

In contrast, this perspective focuses on instructional design rather than primarily on theoretical explanation. Existing frameworks primarily explain why XR influences learning by modeling the relationships among cognitive, affective, and technological factors. However, they offer relatively little guidance on how these constructs should influence instructional design decisions (Makransky and Petersen, 2021; Wee et al., 2025; Yang et al., 2020). Our contribution aims to translate these theoretical insights into concrete design decisions when developing XR learning applications. We therefore reinterpret multimedia learning principles as practical design considerations involving content representation, virtual environment settings, interaction modalities, and learning materials. These design considerations were selected because they represent the primary decisions instructional designers make when creating XR learning experiences, directly shaping how learners perceive information, interact with virtual content, allocate attention, and engage with instructional activities. Rather than introducing another theoretical framework, this work bridges existing XR learning theories and instructional practice by identifying recurring design decisions that directly influence cognitive processing in immersive environments.

2 XR design considerations from a cognitive load perspective

From our perspective, three categories of design decisions are especially important when reinterpreting multimedia learning principles for XR varying across interaction modalities: (1) content representation, (2) environmental design, and (3) the structure of learning materials. These categories are not merely technical choices; they shape how learners allocate attention, manage cognitive load, and engage with instructional content in immersive environments. Grounding these aspects in how working memory processes information and *cognitive load theory*, particularly intrinsic, extraneous, and germane cognitive load, highlights their relevance for designing XR educational applications. Where direct empirical evidence for a specific XR design choice remains limited, we present the following considerations as theoretically informed design propositions rather than established prescriptions; their effects on cognitive load and learning should be validated empirically.

More specifically, the proposed design considerations for XR can be understood as reinterpretations of the principles of multimedia learning. They should not be treated as separate design guidelines. Principles such as signaling, redundancy, spatial and temporal contiguity, segmenting, and modality remain relevant in XR. However, their implementation changes when instructional information is distributed throughout a three-dimensional environment and learners can actively navigate, manipulate, and respond to that information. Table 1 summarizes these relationships and illustrates how established multimedia learning principles inform the XR-specific design considerations developed in this perspective.

2.1 Content representation

In XR, how content is represented is particularly important because users may directly interact with instructional elements. Moreover, because content can surround the learner within the immersive environment, poorly designed representations may introduce unnecessary cognitive demands. Familiarity with

TABLE 1 Relationship between multimedia learning principles and their reinterpretation within the proposed XR design considerations.

Multimedia learning principle	Traditional implication	Reinterpretation for XR design
Multimedia	Learning can benefit from combining words and pictures.	XR extends multimedia representation beyond words and pictures to spatial 2D/3D content, embodied interaction, sound, and potentially haptic information; additional modalities should therefore be included only when they contribute to the learning objective.
Coherence	Extraneous material should be excluded.	Environmental detail, decorative objects, sounds, and other immersive stimuli should be limited when they do not contribute to the instructional task, particularly because XR surrounds the learner with information beyond a bounded screen.
Signaling	Cues should highlight the organization and relevant elements of instructional material.	Color, spatial positioning, sound, highlighting, and other multimodal cues can direct attention toward relevant objects or locations within a three-dimensional environment.
Redundancy	Presenting identical information through unnecessary concurrent formats can impair learning.	XR designers should avoid duplicating instructional information across text, narration, visual cues, or feedback channels when the duplication adds sensory processing without improving comprehension.
Spatial contiguity	Corresponding words and pictures should be presented near one another.	Related instructional information should be spatially associated with the relevant virtual object or location, while accounting for field of view, head movement, occlusion, and the possibility that learners physically navigate the environment.
Temporal contiguity	Corresponding words and pictures should be presented simultaneously rather than successively.	Narration, visual events, haptic cues, and other feedback should be temporally aligned with the learner's actions or the corresponding events in the virtual environment.
Segmenting	Complex multimedia material should be divided into manageable learner-paced segments.	Complex XR activities can be organized into stages or interaction sequences that allow learners to progressively explore content rather than simultaneously processing the full environment and task.
Pre-training	Learners benefit from prior knowledge of key concepts and components.	Pre-training can familiarize learners not only with instructional concepts but also with XR controls, interaction techniques, and spatial conventions before the primary learning activity begins.
Modality	Spoken words can reduce demands on the visual channel when accompanying graphics.	Voice and auditory presentation may reduce competition for visual attention in visually dense XR environments, although auditory information must be coordinated with environmental sounds and other feedback.
Personalization	Conversational language can promote deeper engagement than formal language.	Conversational instruction may be delivered through embodied pedagogical agents or voice-based interfaces, making the social and spatial context of the interaction part of the instructional design.
Voice	Human voice can support learning more effectively than machine-generated voice in some multimedia contexts.	Voice characteristics remain relevant for narration and virtual agents, but XR additionally requires consideration of spatialized audio, agent embodiment, responsiveness, and conversational interaction.
Image	Displaying the speaker's image does not necessarily improve learning.	In XR, the presence and appearance of an embodied virtual instructor or pedagogical agent should be justified by its instructional function because an agent occupies spatial and attentional resources within the learning environment.

content representation may help novice XR users mitigate extraneous load; color can also help keep the focus on the main learning activity, thereby maximizing germane load. Moreover, element positioning can help reduce extraneous load while simplifying intrinsic load.

2.1.1 2D and 3D elements

When visualizing complex phenomena, the appropriate dimensionality depends on the complexity of the phenomenon, the intended interactions, the learning task, and the instructional context. Specifically, the choice between 2D and 3D graphical representations can influence how learners understand the intended concept (see Figure 2). In XR, users likely expect to be surrounded by 3D elements due to the stereoscopic nature of the HMD. However, 2D components (e.g., menus) persist in this environment. In contrast, having 3D interactions, such as gesture-based actions, can serve as an alternative to menus. This immersive layer requires additional time and motor effort. To address possible cognitive overload, designers could embed 2D elements/interactions into the XR experience, which may be perceived as natural/common by

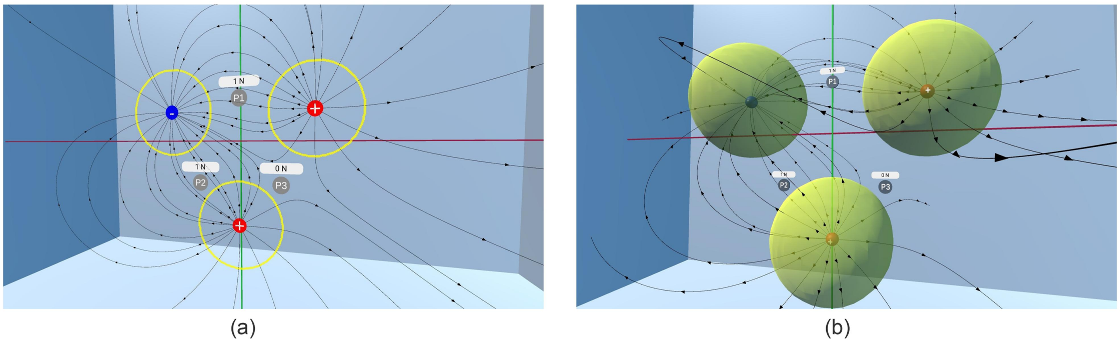

FIGURE 2
Comparison of charged-particle simulations in XR using **(a)** 2D graphical representations and **(b)** immersive 3D visualizations to support spatial understanding of electric field.

novice XR users, and replicate these interactions from traditional desktop-based environments. Thus, we suggest that this approach may reduce the load associated with learning newer interactions (i.e., extraneous), potentially allowing learners to devote more cognitive resources to conceptual understanding.

2.1.2 Color scheme

Color theory can also be applied to XR learning experiences. Color can influence users' perceptions of their overall experience and decisions during interaction with any computer system. We can understand this concept through games, where the player's decision, such as selecting a character, can affect future interactions. Prior work shows that color choices can influence flow, immersion, and perceived competence (Kao and Harrell, 2016). These insights can be extended to XR learning environments, where color choices can either support or hinder cognitive processing.

Strategic use of color may help reduce extraneous cognitive load by clearly distinguishing instructional content from background elements. For instance, in our previous work (Acevedo et al., 2024), we designed a physics learning environment that used primary colors (i.e., blue, red, yellow) for interactive elements and bright tones for background areas to isolate the learning material. Such choices may have influenced participants' sense of self-efficacy by reinforcing focus on core tasks. More broadly, color decisions can affect cognitive processing by guiding attention to relevant information, clarifying distinctions between instructional and noninstructional elements, and supporting engagement with the learning task. In this sense, we propose that color may primarily help reduce extraneous cognitive load by signaling what learners should attend to, while also potentially supporting germane processing when it helps learners organize and integrate relevant information.

2.1.3 Positioning

In XR, we are often immersed in a simulated 3D world that attempts to replicate our real-world environment. This environment can be interactive and engaging, requiring users to move or teleport between locations as they progress through an activity. The placement of learning materials (i.e., slides, simulations, or interactive elements) directly affects attention, engagement, and cognitive processing. While many tasks can be designed for a static position with content displayed in the user's forward view, other XR learning experiences may involve alternative arrangements, such as fixed-screen instructions, spatially distributed activities, or open-world exploration. Similarly, in more passive immersive interactions, the placement of instructional content becomes particularly important, given the limited field of view (FOV) of low-end devices. Belani et al. (2023) highlighted the importance of these choices, showing that users often prefer instructional elements placed directly on objects around the environment. Poorly placed elements risk overwhelming learners or obscuring important information, thereby increasing extraneous cognitive load. Effective positioning requires careful alignment among instructional intent, user guidance, and the need to minimize unnecessary distractions.

2.2 Environment design

The design of the virtual environment plays a crucial role in shaping learners' experiences. While a 2D screen confines attention to the immediate content, a 360° virtual world introduces additional stimuli. Careful consideration is therefore required to balance environmental elements with instructional goals. Similarly, controlling environmental elements can help manage cognitive load by reducing unnecessary stimuli and reinforcing instructional focus.

2.2.1 Background elements

We should clearly distinguish between the learning content (e.g., slides, virtual manipulatives, simulations, or virtual avatars) and the environment (e.g., background elements, the surrounding space, the room's walls, and the horizon). Background features should be selected strategically, as they can strongly influence learners. On the one hand, realistic scenery, such as specialized equipment in a chemistry lab or furnishings in a hospital room, can enhance presence and situate learners within authentic contexts. On the other hand, overly detailed or visually rich backgrounds may divert attention and impose extraneous cognitive load, reducing focus on the intended learning content. This can be an issue in augmented reality (AR) settings where the background is fixed, and content should be presented in a specific format, including color and size, to be noticeable across both environments (Kruijff et al., 2019). This perspective indicates that background realism in XR should be considered as a potential instructional variable rather than merely an aesthetic choice, because environmental detail may support situated learning but may also impose extraneous cognitive load. This proposition warrants empirical examination across different learning tasks and XR environments.

2.2.2 Haptic feedback

In XR, integrating haptic feedback can create a "sense of touch" that makes interactions feel more realistic (Yuksel et al., 2019). For instance, learners can sense how the force exerted by a charged particle varies through differences in haptic feedback intensity or which forces are applied to a glucose molecule on a cell membrane (Webb et al., 2022). This sensory channel can heighten immersion, increase engagement, and support better performance by reinforcing the illusion of physical presence. However, fidelity in haptic design comes at a cost. If feedback is ambiguous or does not align with user expectations, it may prompt additional cognitive effort as learners attempt to interpret its source or meaning. Rather than enhancing the learning task, this could increase extraneous cognitive load as learners must process unclear, unnecessary, or misaligned feedback. It is reasonable to hypothesize that haptic feedback may be most appropriate when it is expected to clarify the learning task or reinforce conceptual understanding; otherwise, it may become an immersive distraction instead of an instructional affordance. The extent to which task-relevant haptic feedback improves learning while avoiding additional cognitive demands remains an empirical question. Although many contemporary HMDs provide basic vibrotactile feedback through controllers, richer or more precise haptic interactions may require specialized peripherals, such as haptic gloves or dedicated input devices.

2.2.3 Sound effects

Sound is another essential component of the virtual environment, contributing to immersion and engagement by reinforcing the sense of "being there." In educational XR, sound can serve as both a feedback mechanism and a contextual enhancer, supporting more realistic and engaging experiences. However, findings suggest that sound does not always significantly influence presence or enjoyment; for instance, XR museum exhibits have shown mixed effects (Kaplan-Rakowski et al., 2023). This highlights the need to carefully evaluate when sound effects meaningfully contribute to learning and when they risk adding unnecessary sensory load. We therefore propose that sound in XR learning is most beneficial when it serves an instructional signaling function, such as guiding attention or providing task-relevant feedback. In contrast, sound that functions primarily as environmental decoration may add sensory load without improving learning. Whether such use reduces extraneous cognitive load should be empirically evaluated across different XR learning contexts.

2.2.4 Voice interactions

Voice interactions, including spoken commands and voice feedback, are a valuable form of interaction in XR environments. By allowing users to control their surroundings through speech rather than physical gestures, they offer a hands-free alternative that can support accessibility and efficiency. For instance, an AI-driven instructor could conduct interactive lessons with real-time question-and-answer exchanges (Zhou and Divekar, 2025). Such interactions may support deeper knowledge exploration and provide opportunities for self-directed learning without the constant presence of a human instructor. However, conversational systems should be designed with attention to potential user fatigue, particularly during extended sessions. Excessive verbal exchanges or poorly tuned voice feedback can overwhelm learners or interrupt concentration. Moreover, speech recognition systems should accurately interpret users' input regardless of accent, pronunciation, or tone; mismatches between user intent and system response can lead to frustration (i.e., extraneous).

2.3 Learning materials

Instructional design for XR requires reconsidering how learning content is presented, sequenced, and assessed within an immersive experience. These decisions can influence multiple dimensions of cognitive load: pre-training may help manage intrinsic load, well-designed interactions and tasks may reduce extraneous load and support germane processing, and appropriately integrated assessments may evaluate learning without unnecessarily disrupting the immersive experience.

2.3.1 Pre-training materials

Traditional media can complement XR by preparing learners before they enter an immersive environment. Because unfamiliar XR technologies may themselves introduce additional cognitive demands, providing an overview or pre-training activity, such as reviewing a 2D infographic, can help learners anticipate the concepts, content, or interactions they will encounter in the virtual environment. Research has demonstrated the benefits of pre-training, which is consistent with Mayer's principles and can provide learners with foundational knowledge before they engage in the primary activity, thereby helping to manage intrinsic cognitive load (Fiorella and Mayer, 2022). Importantly, these materials do not need to be adapted into 3D; instead, they can be delivered through conventional formats before immersion begins, freeing XR for experiential learning.

2.3.2 Interactions and tasks

Designing how learners interact with instructional content is a central challenge in XR learning experiences. Overly complex interactions may increase cognitive demands and potentially interfere with learning by shifting attention from the instructional content to operating the XR system. Interaction complexity may therefore influence whether XR provides an instructional advantage over less immersive alternatives. Interaction techniques should be as intuitive and accessible as possible so that learners can devote their cognitive resources to the learning content, not to operating the system. In line with Mayer's principle, we cannot ignore the embodiment principle in task completion (Fiorella, 2021). The embodiment principle also suggests that observing another person perform a task may support learning by providing a model for subsequent action (Fiorella, 2021). XR can extend this approach by allowing learners to observe, imitate, or collaboratively perform embodied tasks within the same environment.

2.3.3 Assessment

Assessment strategies must also be carefully adapted for XR environments. While in-XR testing can help create a self-contained learning ecosystem, traditional multiple-choice or open-ended formats may not translate seamlessly into XR. The familiarity of web-based assessments suggests that some evaluations remain more effective in 2D. However, when integrating assessments into XR, simple interfaces, such as cursor or ray-casting selection, can reduce complexity while preserving immersion. The assessment format, such as multiple-choice or open-ended questions, can also affect students' performance, highlighting the importance of carefully selecting both the assessment format and its implementation (Belga et al., 2022). Designers should evaluate whether there is a potential for distraction or increased cognitive demand.

3 Discussion

Prior research has shown that immersive XR can support learning and has offered general guidance for its effective use (Makransky and Petersen, 2021). However, much of this work applies existing frameworks without fully addressing the distinctive cognitive and interaction demands of XR environments. Although some studies have examined specific XR design choices (Acevedo et al., 2024; Webb et al., 2022; Belani et al., 2023), more explicit design guidance is needed to help researchers and designers adapt multimedia learning principles to XR environments while accounting for cognitive load. From our perspective, design considerations related to content representation, virtual environment settings, and the structuring of learning materials are critical factors shaping engagement and learning effectiveness.

Beyond summarizing these design considerations, this perspective highlights important theoretical implications. Specifically, we argue that traditional multimedia learning principles are insufficient when directly applied to XR environments. They do not fully account for embodiment, spatial interaction, or environmental complexity, and these factors vary across XR systems. An additional consideration is that XR should not be treated as a single technological platform. The cognitive and interaction demands of XR vary substantially across devices, ranging from smartphone-based viewers with gaze interaction to standalone headsets with six degrees of freedom, hand tracking, haptic feedback, and spatial computing capabilities. For design purposes, these systems can be broadly considered in three categories: smartphone-based or low-interactivity immersive displays, fully interactive VR headsets, and XR systems that combine digital content with the physical environment. Each category introduces different constraints on the transferability of multimedia learning principles.

Smartphone-based viewers and other low-interactivity systems typically provide limited tracking and input capabilities. In these environments, instructional design may rely more heavily on gaze, passive observation, and simple selection mechanisms. Such limitations may constrain embodiment and agency, but they may also reduce the cognitive demands associated with learning complex interaction techniques. Fully interactive VR headsets, by contrast, can support spatial movement, six-degree-of-freedom tracking, hand- or controller-based manipulation, and haptic feedback. These affordances create greater opportunities for embodied and experiential learning but also introduce additional motor, spatial, and interaction demands that may contribute to extraneous cognitive load. XR systems introduce a different design challenge because virtual instructional information must coexist with stimuli from the physical environment. In these systems, signaling, positioning, visual contrast, and management of competing physical and digital information may become particularly important.

These differences influence how learners access information, interact with virtual content, and distribute cognitive resources. For example, reduced FOVs may increase split attention by requiring users to search for instructional content. Limited interaction capabilities may reduce opportunities for embodied learning, but they may also simplify interaction complexity. Thus, a design proposition may function differently across device categories. For example, spatially distributing instructional information may support exploration in a fully tracked VR environment but increase search demands on a device with a restricted field of view. Similarly, haptic signaling is only relevant when the platform provides an appropriate feedback mechanism, while principles concerning visual contrast and the separation of instructional content from background information may become especially important in XR environments where designers have less control over the user's surroundings.

Consequently, the application of multimedia learning principles in XR should be viewed as hardware-dependent, as design strategies that are effective on fully interactive headsets may require modification when implemented on devices with constrained interaction, tracking, or display capabilities. Accordingly, the proposed design considerations should be evaluated in relation to the display, tracking, interaction, and feedback capabilities of the target platform rather than assumed to transfer uniformly across XR devices. These hardware differences may partly explain mixed findings in prior research applying multimedia learning principles to XR.

Importantly, the relevance of these design considerations is likely to depend on several contextual factors. XR platform characteristics may alter both cognitive demands and available interaction possibilities. For example, smartphone-based viewers, augmented reality displays, and fully interactive headsets differ in field of view, tracking fidelity, input modalities, and feedback capabilities. Learners' prior experience with XR may also influence how much cognitive effort is devoted to understanding the interface rather than the instructional content, with novice users potentially requiring greater familiarization or pre-training. In addition, disciplinary domain may shape which XR affordances are most instructionally meaningful, while task complexity may moderate the usefulness of immersive features and the need for scaffolding, simplified interactions, or reduced environmental detail. These factors suggest that the proposed design considerations should not be interpreted in isolation from the learner, task, domain, and technological context in which they are implemented.

Building on these contextual considerations, the following recommendations synthesize the preceding discussion into a set of testable design propositions. They are not intended as a finalized framework. Instead, they identify practical priorities for researchers, instructional designers, and XR developers seeking to reinterpret multimedia learning principles for immersive learning environments. Collectively, the propositions address content representation, environmental design, interaction modalities, and learning-material structure through the lens of cognitive load and XR-specific affordances:

- Consider using 3D elements for realistic scene fidelity and 2D elements for clear navigation and menus.
- A simplified, cohesive color palette may help support focus and reduce cognitive load.
- Positioning content strategically on key objects or within the user's primary field of view may help guide attention.
- Backgrounds should be examined for their potential to enhance presence without distracting from instructional content.
- Haptic cues may be beneficial when they are aligned with learning objectives and when they meaningfully enrich multimodal experiences.
- Sound may be most effective when used selectively to reinforce context while avoiding unnecessary distractions.
- Voice commands and voice feedback may support natural, AI-driven dialogue with the learning content when they reduce rather than add interaction demands.
- Brief pre-training may help build familiarity and ease cognitive processing.
- Interactions may be more effective when they are intuitive and directly aligned with learning objectives.
- Simple in-VR assessments may offer a way to measure learning without breaking immersion, although their effectiveness should be compared with conventional assessment formats.

Because these recommendations are intentionally concise, their implementation requires consideration of the potential benefits, risks, and contextual constraints associated with each design choice. Table 2 therefore elaborates on the propositions by linking each one to its expected cognitive benefit, potential risk, and conditions of applicability. In doing so, the table distinguishes considerations that may apply broadly across XR systems from those whose effectiveness depends on particular learner, task, instructional, or hardware conditions.

While this perspective outlines key design considerations for XR-based learning, it is important to acknowledge its limitations. The proposed guidelines are conceptual and have not been empirically validated across diverse learning contexts, user populations, or XR platforms. Their effectiveness may also vary with learners' prior experience with XR, task complexity, educational domain, and XR hardware capabilities. These factors may also interact with one another. For example, an interaction technique that is intuitive for an experienced XR user may impose substantial extraneous cognitive load on a novice, while the same level of environmental realism may support situated learning in one disciplinary domain but distract from abstract conceptual learning in another. Similarly, complex learning tasks may require stronger signaling, segmentation, or pre-training than simpler tasks, particularly when learners must simultaneously manage unfamiliar interaction techniques. Design choices that are appropriate for high-end standalone headsets may not directly translate to lower-cost smartphone-based viewers due to differences in field of view, tracking fidelity, interaction techniques, and available feedback modalities. Accordingly, the proposed design considerations should be understood as context-sensitive propositions whose effectiveness may depend on the interaction among platform capabilities, learner experience, learner characteristics, educational context, instructional objectives, disciplinary goals, and task demands.

Future research should move beyond applying traditional multimedia learning principles directly to XR and empirically examine how these principles should be adapted across different immersive contexts and hardware configurations. In particular, constructs such as embodiment, presence, agency, spatial interaction, and environmental complexity should be examined as factors that may shape how learners allocate cognitive resources in immersive environments. Developing such frameworks would help the research community better explain mixed findings in XR learning research and provide a stronger theoretical basis for future design principles. In parallel, empirical studies should test how these factors influence cognitive load, learning outcomes, and user experience across different XR settings.

Beyond cognitive and instructional considerations, XR learning experiences must also account for accessibility, hardware limitations, and diverse learner needs. Factors such as motion sickness, visual fatigue, interaction discomfort, and limited familiarity with XR technologies can negatively affect

TABLE 2 Proposed XR design propositions, including their expected cognitive benefits, potential risks, and conditions of applicability.

Design consideration	Expected cognitive benefit	Potential risk	Conditions of applicability
2D/3D representation	May reduce extraneous processing when representation dimensionality matches the instructional task; 3D may support spatial understanding and embodiment.	Unnecessary 3D complexity may increase visual, spatial, and interaction demands.	3D may be most useful when depth, spatial relationships, or embodied manipulation are relevant; familiar 2D elements may be preferable for menus, instructions, or navigation.
Color scheme	May guide attention, distinguish relevant information, and reduce processing of noninstructional elements.	Excessive, inconsistent, or poorly contrasted colors may compete for attention or obscure relevant information.	Most useful when color has a clear signaling function and sufficient contrast is maintained across the virtual and, in XR, physical environment.
Positioning	May direct attention and reduce visual search by placing instructional information near relevant objects or within the learner's primary field of view.	Poorly positioned or widely distributed content may increase split attention and search demands.	Depends on field of view, display type, task requirements, and whether exploration of the surrounding environment is itself part of the learning objective.
Background elements	May support contextualization, situated learning, and presence when environmental information is relevant to the task.	Excessive realism or irrelevant environmental detail may increase extraneous cognitive load and distraction.	Environmental fidelity may be most beneficial when contextual cues contribute directly to learning; simpler environments may be preferable for abstract or cognitively demanding tasks.
Haptic feedback	May reinforce conceptual information through an additional sensory channel and support embodied interaction.	Ambiguous, unnecessary, or poorly synchronized feedback may require additional interpretation and increase extraneous processing.	Most applicable when tactile or force-related information is relevant to the learning objective and appropriate haptic hardware is available.
Sound effects	May provide signaling, feedback, or contextual information without requiring additional visual attention.	Decorative or competing sounds may add unnecessary sensory information and distract from instructional content.	Most useful when sounds communicate task-relevant events, directional cues, or feedback and do not compete with narration or other important auditory information.
Voice interactions	May reduce reliance on menus or physical input and support natural interaction, accessibility, and learner-directed inquiry.	Recognition errors, lengthy exchanges, or irrelevant responses may interrupt concentration and increase extraneous cognitive load.	Most applicable when speech provides a simpler interaction mechanism than manual input and when recognition and system responses are sufficiently accurate and task relevant.
Pre-training	May reduce intrinsic processing demands by familiarizing learners with concepts, controls, and task procedures before immersion.	Excessive pre-training may increase instructional time or duplicate information unnecessarily.	May be particularly useful for novice XR users, unfamiliar interaction techniques, or tasks involving complex concepts and procedures.
Interactions and tasks	May support germane processing and embodiment when actions are intuitive and directly related to the learning objective.	Complex or unfamiliar interaction techniques may shift attention from learning content toward operating the system.	Interaction complexity should reflect learner experience, task complexity, accessibility needs, and the input capabilities of the target XR platform.
Assessment	May maintain continuity of the learning experience and permit assessment within the context in which learning occurs.	Complex in-XR response mechanisms may introduce construct-irrelevant cognitive or motor demands.	In-XR assessment may be appropriate when contextual or performance-based evaluation is important; conventional 2D assessment may remain preferable when immersion offers no assessment benefit.

These propositions are theoretically informed and require further empirical validation across learning contexts and XR platforms.

learners' experiences and cognitive processing. In addition, hardware constraints, including tracking inaccuracies, limited field of view, and device comfort, may influence usability and learning outcomes. Future XR design frameworks should therefore integrate accessibility, comfort, and inclusiveness alongside immersion and engagement.

In conclusion, this perspective bridges multimedia learning theory and XR instructional design. It provides a foundation for future empirical research. It can also guide the development of more effective, evidence-informed immersive learning environments.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

PA: Writing – original draft, Writing – review & editing. AM: Conceptualization, Supervision, Writing – review & editing. BB: Writing – review & editing, Conceptualization, Supervision. CM: Writing – review & editing, Conceptualization, Supervision, Writing – original draft, Project administration.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author CM declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. Minimally for grammar checking and for fewer edits using AI-assisted tools, including Grammarly. This tool was employed solely to assist the authors in language editing. The AI was not used to generate any original research ideas or interpretations.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Acevedo, P., Magana, A. J., Walsh, Y., Will, H., Benes, B., and Mousas, C. (2024). Embodied immersive virtual reality to enhance the conceptual understanding of charged particles: a qualitative study. *Comput. Educ. X Real.* 5, 100075. doi: 10.1016/j.cexr.2024.100075

Belani, M., Singh, H. V., Parnami, A., and Singh, P. (2023). "Investigating spatial representation of learning content in virtual reality learning environments," in *2023 IEEE conference virtual reality and 3D user interfaces (VR)* (IEEE). doi: 10.1109/vr55154.2023.00019

Belga, J., Do, T. D., Ghamandi, R., McMahan, R. P., and LaViola, J. J. (2022). "Carousel: improving the accuracy of virtual reality assessments for inspection training tasks," in *Proceedings of the 28th ACM symposium on virtual reality software and technology* (ACM), VRST '22. doi: 10.1145/3562939.3565618

Çeken, B., and Taşkın, N. (2022). Multimedia learning principles in different learning environments: a systematic review. *Smart Learn. Environ.* 9, 19. doi: 10.1186/s40561-022-00200-2

Fiorella, L., and Mayer, R. E. (2021). Principles based on social cues in multimedia learning: personalization, voice, image, and embodiment principles. in *The Cambridge Handbook of Multimedia Learning,* eds. R. E. Mayer and L. Fiorella, Cambridge: Cambridge University Press (pp. 277–285). doi: 10.1017/9781108894333.030

Fiorella, L., and Mayer, R. E. (2022). Principles based on social cues in multimedia learning: personalization, voice, image, and embodiment principles. *Camb. Handb. Multimed. Learn.* 3, 296–303.

Kao, D., and Harrell, D. F. (2016). "Exploring the impact of avatar color on game experience in educational games," in *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems* (ACM), CHI'16, 1896–1905. doi: 10.1145/2851581.2892281

Kaplan-Rakowski, R., Cockerham, D., and Ferdig, R. E. (2024). The impact of sound and immersive experience on learners when using virtual reality and tablet: a mixed-method study. *Br. J. Educ. Technol.* 55, 1560–1582. doi: 10.1111/bjet.13417

Kruijff, E., Orlosky, J., Kishishita, N., Trepkowski, C., and Kiyokawa, K. (2019). The influence of label design on search performance and noticeability in wide field of view augmented reality displays. *IEEE Trans. Vis. Comput. Graph.* 25, 2821–2837. doi: 10.1109/TVCG.2018.2854737

Makransky, G., and Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. *Educ. Psychol. Rev.* 33, 937–958. doi: 10.1007/s10648-020-09586-2

Mayer, R. E. (2009). *Multimedia Learning* (3rd ed.). Cambridge: Cambridge University Press. doi: 10.1017/cbo9780511811678

Webb, M., Tracey, M., Harwin, W., Tokatli, O., Hwang, F., and Johnson, R., et al. (2022). Haptic-enabled collaborative learning in virtual reality for schools. *Educ. Inf. Technol.* 27, 937–960. doi: 10.1007/s10639-021-10639-4

Wee, C., Wang, L. Y. K., and Ong, H. F. (2025). TeachVR: an immersive virtual reality framework for computational thinking based on student preferences. *ACM Trans. Comput. Educ.* 25, 1–36. doi: 10.1145/3717838

Yang, K., Zhou, X., and Radu, I. (2020). Data from: Xr-ed framework: designing instruction-driven andlearner-centered extended reality systems for education. doi: 10.48550/ARXIV.2010.13779

Yuksel, T., Walsh, Y., Magana, A. J., Nova, N., Krs, V., and Ngambeki, I., et al. (2019). Visuohaptic experiments: exploring the effects of visual and haptic feedback on students' learning of friction concepts. *Comput. Appl. Eng. Educ.* 27, 1376–1401. doi: 10.1002/cae.22157

Zhou, Y., and Divekar, R. (2025). Immersive, task-based language learning through XR and AI: from design thinking to deployment. *TechTrends* 69, 427–446. doi: 10.1007/s11528-025-01048-2