International Society of the Learning Sciences

Interactivity in Extended Reality: Towards A Hierarchical Model for Defining and Designing Learning Experiences

Gregory McGowin, University of Central Florida, gregory.mcgowin@ucf.edu
Stephen M. Fiore, University of Central Florida, sfiore@ucf.edu

Abstract: Extended realities (XR), encompassing virtual , augmented , and mixed reality (MR), offer transformative learning opportunities through interactive experiences. However, the concept of interactivity within XR remains ambiguously defined. This work reviews interdisciplinary definitions of interactivity and constructs a conceptual definition, operationalized using Schulmeister's (2003) hierarchical model. From this, the paper presents a taxonomy of interactivity, clarifies its role as a key learning affordance, and offers a framework for analyzing and designing XR experiences. By exploring XR's unique capabilities for learning, this work guides the development of immersive environments and proposes recommendations for advancing the field. These insights set the stage for future empirical studies examining the relationship between interactivity, affordances, and learning outcomes, ultimately helping developers and educators maximize XR's potential for effective learning.

Introduction

The affordability and prevalence of Extended Reality (XR) technology, encompassing virtual, augmented, and mixed realities has led to an increased interest in how this technology can be used for learning in both training and education. If properly implemented, XR can be a potent training tool that simulates real-world experiences, provides learners with an interactive multimedia environment, and allows them to practice complex skills and procedures through experiential learning (Kaplan et al., 2021; Radianti et al., 2021). In addition, XR offers a safe and cost-effective way to engage in learning activities that may be too dangerous or expensive to perform in the real world (Kaplan et al., 2021). Studies suggest that XR can improve learning outcomes and provide at least comparable results to traditional training methods (Kaplan et al., 2021; Radianti et al., 2021). However, exploring the relationship between XR and learning is a complex, multifaceted endeavor, with scholars highlighting several key affordances of the technology, including immersion, presence, agency, embodiment, and interactivity (Johnson-Glenberg, 2018; Klingenberg et al, 2024; Makranksky & Petersen, 2021, McGowin et al., 2023, McGowin & Fiore, 2024). McGowin and colleagues recently developed a broad-based theory of I-VR learning affordances, grounded in the cognitive sciences framework of 4E cognition. For this paper, we draw upon their prior research and concentrate on the relationship between interactivity and extended cognition. Interactivity is frequently cited as an important feature in existing literature on XR and learning. However, in the XR learning literature, it is generally underdefined and underexplored (e.g., Andreasen et al., 2021, Petersen & Makranksy, 2021; Klingenberg et la., 2024). This paper reviews various definitions of interactivity across multiple domains, presents a taxonomy of interactivity derived from existing literature, and illustrates, through examples, how interactivity can be applied in the diverse technologies within extended reality. Additionally, it offers a set of recommendations to further advance research and development in the learning sciences.

Interactivity

Over two decades since McMillan noted, "interactivity means different things to different people in different contexts" (McMillan, 2002, as cited in Moreno & Mayer, 2007, p. 163), the field still struggles with a coherent and concise conceptualization, creating multiple definitions within and across domains. For context, Powell and Leary (2021) conducted a content analysis of interactivity (in the context of learner-content interaction) in which they examined 107 articles to identify how constructs of interaction are used. They found that from the 107 articles, 113 separate measurement tools were used for interactivity, indicating the substantial range of measurement tools related to interaction. They further noted that most of the tools were not generalizable and were unique to each study context (Powell, & Leary, 2021).

Early definitions from the communications literature described interactivity as "an expression of the extent that in a given series of communication exchanges, any third (or later) transmission (or message) is related to the degree to which previous exchanges referred to even earlier transmissions" (Rafaeli, 1988, p. 111). Steuer (1992) expanded on this idea, placing interactivity in the context of a mediated environment (ME) and defining it as "the extent to which users...can [modify] the form and content of a mediated environment in real time" (p. 14), consolidating Sheridan's (1992) and Zeltzer's (1992) models into a single dimension centered on user control.

 © ISLS

Steuer (1992) outlines three aspects of interactivity, namely speed, range, and mapping. Speed relates to the concern of real-time interactions, range covers the actions a user can perform, both in temporal and spatial ordering, as well as in intensity (e.g., loudness, brightness) of the ME. Mapping captures how user interactions are 'mapped' onto the virtual environment (VE), on a continuum from completely arbitrary, to completely natural, or also referred to in some literature as 'congruence' in gesture or sensory-motor representation (e.g., Johnson-Glenberg, 2018; Kilteni et al., 2012).

In terms of multimodal learning environments, Moreno and Mayer (2007) define interactivity in relation to the responsiveness of the system to the learners actions (i.e., multidirectional communication between the user and the mediated environment) with the purpose of fostering learning. Hence, a single direction (i.e., passive) multimedia learning experience presented in a linear fashion would not qualify as 'interactive'. From this, they provide five forms of interactivity for mediated learning environments: dialoguing, controlling, manipulating, searching, and navigating. Dialoguing is the question and answer, or feedback, provided from the system (i.e., a learner can ask a question about a topic and receive a predetermined response from a virtual agent). Controlling is relegated to pace and order that information is presented to the learner, somewhat aligning to Makranksy and Petersen's (2021) definition of agency. Manipulating encapsulates both setting simulation parameters and control of viewpoints (i.e., zoom in/out, rotation of 3D objects). Searching allows a learner to find and select information. Navigating encompasses moving through different menus to display different content.

When conceptualizing *varieties* of interaction, Wu et al. (2005) describe a set of differences between cognitively and perceptually mediated models, with the underlying distinction about mode of control. Here, we can use Steuer's (1992) continuum of arbitrary to natural user input as a frame of reference. Cognitively mediated models are controlled via arbitrary interactions (e.g., mouse and keyboard) whereas perceptually mediated models rely on 'natural' or visuomotor control, such as physical/direct manipulation (Barret et al., 2015).

Schwier (1992) has a three-pronged approach to interactivity, labeling the components reactive, proactive, and mutual. Reactive is a two-way interaction between the learner and the ME, such that the system displays a stimulus to the learner in the forms of questions or information, and the learner 'reacts' to said stimuli, by selecting an answer or navigating to a new area of content. Proactive focuses on the generative processes of learning that go beyond just selecting or navigating and is focused on "elaborations beyond designer-imposed limits" (p.5). Finally, the highest level of interaction is mutual interaction. This level is the home to both artificial intelligence (AI) and virtual (and by extension, extended) reality (Schweir, 1992). Here the learner and system have a recursive and mutually adaptive (e.g., system in systems) relationship, with the main point being that the system/learner interaction is capable of essentially unlimited dialoguing (i.e., an unlimited conversation between the human agent and the machine agent). These levels are not mutually exclusive and often work in tandem, but they are hierarchical. Schwier (1992) outlines how these concepts are interconnected stating "when learners generate new questions and approaches (proactive) they can…be used by the system to formulate new conversations (mutual). Similarly, when learners generate their own strategies (proactive) they are responding to existing stimuli at a sophisticated level (reactive)" (p. 6).

Schulmeister (2003) offers a different perspective with his taxonomy of interactivity, presenting a six-level continuum. This framework ranges from minimal interactivity in a mediated environment (Level I) to constructing content and receiving intelligent feedback (Level VI). To contextualize this continuum at its first level, learners can pause a video or select a text prompt. At the other end of the continuum is the ability to create or manipulate the medium and receive meaningful feedback (See Appendix 1 for the continuum of interactivity and examples of both traditional and XR learning environments). Despite predating much of the theoretical and empirical work, Schulmeister (2003) still provides the most comprehensive hierarchical model of interactivity, and the one that encompasses many elements of the others. Thus, we base our operational definition of interactivity on Schulmeister's (2003) work. With this, we synthesize other definitions to distill them into both conceptual and operational constructs. Our conceptual definition for interactivity is as follows: Interactivity is the multidirectional exchange of related information between two or more agents (i.e., persons, system, virtual agents) (Rafaeli, 1988), which can be characterized by the degree of modification, malleability, and recursive feedback between agents in a mediated environment (Schwier; 1992; Steuer,1992; McGowin et al., 2023). This definition encompasses multiple levels of interactivity, from a person reading a book or viewing an animation to a recursive dynamical system providing virtually unlimited dialogue or feedback. For our operational definition of interactivity, we build off the works of Schulmeister (2003), Stalph and Heravi (2021), and McGowin et al., (2023) (see Appendix 1 for additional details). Next we provide a more in-depth examination of the continuum of interactivity including concrete examples across various XR learning environments.

Level I: Viewing and receiving content

To illustrate, consider reading a book: the reader gains information and interacts minimally, mainly by turning pages or choosing a section. Similarly, this level of interactivity is comparable to watching videos where the viewer either passively watches or actively skips to different parts of the video (i.e., skip to location, but no pause function). In an XR environment, such as VR, Level I could manifest as passively viewing immersive videos or exploring 360-degree photo environments. Here, interaction is primarily limited to navigating one's viewpoint around the virtual space and occasionally selecting specific locations or sequences, such as deciding the order in which to view exhibits in a virtual museum.

Level II: Watching and receiving multiple representations

This level can include multiple streams of information to elaborate on what the learner is consuming, such as a GIF providing supplemental information about a process in which the learner is engaged. In XR, this level of interactivity (II) allows users to engage with multiple types of media or informational overlays within a VE. They can control not only the timing of their interactions, (e.g., pause, skip, select), but also customize the placement and display of information streams. For example, users might engage an AR display to view additional architectural details on one side of their visual field while simultaneously accessing historical facts or on the other, possibly enhancing their sense of presence and learning experience. Note that the component representation does not allow anything beyond watching/listening/reading and only serve for illustration or information.

Level III: Varying the form of representation

At this level, users control how information is presented without altering its content. They can rotate, scale, or cross-section objects, fostering active engagement. For example, a user may switch between first- and third-person views of a carbon molecule, but this does not change its structure or properties (e.g., it remains a carbon molecule and does not become another structure such as graphene). In XR, this interactivity (III) enables users to manipulate objects in three dimensions, enhancing understanding through direct interaction. Users can pick up, rotate, and resize virtual artifacts, examining them from multiple angles without changing their properties. For example, in an MR/AR archaeology application, a student could view a digital artifact overlaid on a physical table, rotate it for detailed examination, or resize it to match a real-world reference. These interactions mimic real-world exploration, reinforcing understanding and making abstract concepts more tangible and concrete.

Level IV: Manipulating the component content

Learners can generate non-predefined outputs by selecting relevant inputs and entering them into the system. This is the first-time users are ‘creating’ content through their interactions with the system. For example, in an educational software application, students might input different variables into a simulation to observe how changes in these variables affect the outcome of an experiment, creating unique scenarios based on their choices. For XR, level IV interactivity empowers users to actively manipulate content within the virtual environment. This level involves more than simple object manipulation; it includes the ability to alter environmental conditions, change the course of events, or interact with complex systems that react dynamically to user decisions. For example, users might adjust lighting and weather conditions to see how they affect a virtual ecosystem, or they could solve problems that alter the storyline in an interactive narrative. These changes directly impact the virtual world, providing immediate feedback on their actions (i.e., users ‘see’ the output of the manipulations).

Level V: Constructing the object or representations from user input

At this level, learners are able to interact with the system to create new objects with the help of a comprehensive toolset developed specifically for this purpose (i.e., created by the developer of the system). For instance, platforms like Roblox, Minecraft, and Second Life provide these types of opportunities by allowing users to design and construct intricate virtual structures and environments from scratch via an expansive toolset. In XR, level V interactivity is distinguished by the extensive creative freedoms provided by advanced toolsets. These tools could allow users to design, construct, and even program objects or environments from scratch. For instance, users could design the architecture of an entire virtual building, sculpt intricate artworks, or engineer complex machinery, all within the confines of a virtual or mixed space. This level of interactivity not only fosters creativity but could also enhance practical skills such as spatial reasoning, architectural planning, and artistic expression.

Level VI: Constructing the Object or Contents of the Representation and Receiving Intelligent Feedback from the System through Manipulative Action

This highest level of interactivity integrates AI—whether real AI or algorithmic feedback—to enhance learning across all lower interactivity levels. Adaptive learning platforms exemplify this by analyzing student interactions, adjusting task difficulty, recommending personalized learning paths, and delivering targeted feedback. In XR,

Level VI interactivity similarly leverages AI to refine skills and modify learner behaviors within complex simulations. For instance, an AR or VR surgical training program could analyze a trainee's technique, provide real-time corrective guidance, and dynamically adjust task difficulty. This ensures personalized, contextually relevant feedback that directly informs and enhances the learner's actions.

Appendix 1 provides a series of concrete examples of how interactivity can be applied to both XR and more traditional learning environments (i.e., non-XR).

Recommendations

R1 - Metrics for interactivity in XR

Building on Schulmeister's (2003) hierarchical model of interactivity, future research should focus on creating standardized tools and methodologies for quantitatively and qualitatively assessing interactivity across different levels in XR environments. While Schulmeister's work outlines the conceptual framework for interactivity, there is a gap in how these levels can be practically evaluated. Developing assessment metrics would bridge this gap, enabling researchers and designers to evaluate how varying degrees of interactivity (e.g., passive viewing vs. manipulative actions) are integrated into XR experiences and their impacts on learning outcomes. These metrics would allow for a more systematic exploration of the relationship between interactivity and learner engagement, supporting the advancement of XR research and the effective design of immersive learning environments. By further operationalizing Schulmeister's framework, such tools would provide a consistent approach for evaluating interactivity in XR across various applications and learning contexts.

R2 - Interactivity and XR affordances

Interactivity does not exist in isolation but interacts dynamically with other XR affordances such as immersion, agency, and embodiment. Future research should investigate these interactions to identify how they collectively influence learning outcomes. For instance, how does the degree of interactivity impact a learner's sense of presence, or agency, and how do these affordances enhance (or detract) from the learning experience? Studies should focus on identifying the optimal balance of these factors for different types of learning tasks, such as procedural training versus conceptual or declarative understanding. Experimental designs could manipulate each affordance with corresponding levels of interactivity to begin to isolate their individual and combined effects on engagement, knowledge transfer, and skill acquisition. This line of inquiry will provide deeper insights into how XR environments can be fine-tuned for specific pedagogical and andragogical goals.

R3 – CPD and Interactivity

As XR environments increase in interactivity, they offer immersive learning experiences that engage users in novel ways. However, the complexity introduced by greater interactivity may also increase Cognitive Processing Demand (CPD), an umbrella term that encompasses various dimensions, including cognitive load and mental task load, and refers to the cognitive/mental demands placed on learners, which may hinder the learning process (McGowin et al., 2024). At a high-level, CPD theories suggest that excessive cognitive demands diminish the ability to process and retain information effectively. Thus, future research should explore how the increasing interactivity of XR environments affects CPD factors. For example, how do task complexity and learner expertise intersect with XR interactivity to influence learning outcomes? An open question is whether adding more interactive elements in XR, especially for novice learners, leads to excessive cognitive load that detracts from learning, or at the other end of the expertise spectrum, can expert learners effectively manage these increased demands without negative consequences. Additionally, research should focus on adaptive strategies for managing CPD in XR environments (e.g., dynamic scaffolding and fading) to balance task challenge with learner ability. By examining how interactivity levels interact with CPD, we can gain insight into 'optimal' levels of interactivity for diverse learning contexts and learner characteristics.

Conclusion

Schulmeister's (2003) levels of interactivity outline a progression from passive content consumption to active manipulation and intelligent feedback. This paper demonstrates how XR can enhance learning by leveraging these interactivity levels, with future research needed to standardize interactivity metrics, investigate the relationship between interactivity and XR affordances, explore AI-driven feedback, and examine how cognitive processing demand interacts with different levels of interactivity. Adapting frameworks like Schulmeister's can help establish a shared understanding of interactivity in XR, enabling more effective collaboration, systematic design, and impactful learning experiences.

References

Andreasen, N. K., Baceviciute, S., Pande, P., & Makransky, G. (2019). Virtual Reality Instruction Followed by Enactment Can Increase Procedural Knowledge in a Science Lesson. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 840–841. https://doi.org/10.1109/VR.2019.8797755

Johnson-Glenberg, M. C. (2018). Immersive VR and Education: Embodied Design Principles That Include Gesture and Hand Controls. *Frontiers in Robotics and AI*, *5*, 1-19.

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. *Human Factors*, *63*(4), 706–726. https://doi.org/10.1177/0018720820904229

Kilteni, K., Normand, J.-M., Sanchez-Vives, M. V., & Slater, M. (2012). Extending Body Space in Immersive Virtual Reality: A Very Long Arm Illusion. PLoS ONE, 7(7), e40867.

Klingenberg, S., Bosse, R., Mayer, R. E., & Makransky, G. (2024). Does Embodiment in Virtual Reality Boost Learning Transfer? Testing an Immersion-Interactivity Framework. *Educational Psychology Review*, *36*(4), 116.

Macrine, S. L., & Fugate, J. M. B. (2022). Embodied Cognition and Its Educational Significance. In S. L. Macrine & J. M. B. Fugate (Eds.), Movement Matters: How Embodied Cognition Informs Teaching and Learning (p. 0). The MIT Press. https://doi.org/10.7551/mitpress/13593.003.0006

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*, *33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

McMillan, S. J. (2002). Exploring models of interactivity from multiple research traditions: Users, documents, and systems. *Handbook of new media*, *2*, 163-182.

McGowin M., Fiore S. M., Oden K. (2023). Towards a theory of learning in immersive virtual reality: Designing learning affordances with embodied, enactive, embedded, and extended cognition. In Cherner T., Fegely A. (Eds.), *Bridging the XR technology-to-practice gap: Methods and strategies for blending extended realities into classroom instruction* (pp. 35–53). Association for the Advancement of Computing in Education.

McGowin, G., & Fiore, S. M. (2024). Mind the Gap! Advancing Immersion in Virtual Reality—Factors, Measurement, and Research Opportunities. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (p. 10711813241278833). Sage CA: Los Angeles, CA: SAGE Publications.

McGowin, G., Sonnenfeld, N. A., & Fiore, S. M. (2024). Navigating Cognitive Demand in Virtual Reality: Implications for Education and Training. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (p. 10711813241278286). Sage CA: Los Angeles, CA: SAGE Publications.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments: Special issue on interactive learning environments: Contemporary issues and trends. *Educational psychology review*, *19*, 309-326.Newen, A., De Bruin, L., & Gallagher, S. (2018). 4E cognition: Historical roots, key concepts, and central issues. The Oxford handbook of 4E cognition, 2-16.

Powell, S. T., & Leary, H. (2021). Measuring learner–content interaction in digitally augmented learning experiences. *Distance Education*, *42*(4), 520-546.

Rafaeli, S. (1988). Interactivity: From new media to communication. *Sage Annual Review of Communication Research: Advancing Communication Science, 16*, 110–134.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 1-29. https://doi.org/10.1016/j.compedu.2019.103778

Schulmeister, R. (2003). Taxonomy of multimedia component interactivity. A contribution to the current metadata debate. *Studies in Communication Sciences. Studi di scienze della communicazione*, *3*(1), 61-80.

Stalph, F., & Heravi, B. (2021). Exploring data visualisations: An analytical framework based on dimensional components of data artefacts in journalism. *Digital journalism*, 1-23.

Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, *42*(4), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.

Schwier, R. A. (1992). *A taxonomy of interaction for instructional multimedia* [Paper presentation]. Annual Conference of the Association for Media and Technology, Vancouver, British Columbia.

Wu, B., Klatzky, R. L., Shelton, D., & Stetten, G. D. (2005). Psychophysical evaluation of in-situ ultrasound visualization. *IEEE Transactions on Visualization and Computer Graphics*, *11*(6), 684–693.

Appendix 1

Levels of Interactivity with Examples of Both Traditional and XR.; adapted from Schulmeister (2003), Stalph and Heravi (2021), and McGowin et al., (2023).

Level of Interactivity	Description	Traditional (Non-XR) Example	XR Example
Level I	**Viewing content** (i.e., watch, read, listen). Minimal to no interaction.	Reading a book or watching a video with minimal interaction.	In a museum experience, users passively view 360-degree recordings of historical events or virtually tour exhibits with limited interactivity beyond looking around.
Level II	**Watching and receiving multiple representations** – with or without control over pacing (i.e., multiple information streams).	Viewing a GIF or supplemental information alongside the main content.	In an educational sim, users can display supplemental content like diagrams or videos alongside the main virtual or real environment, controlling when and where it appears in their view.
Level III	**Varying the form of representation** – (e.g., rotating, scaling, or cross-sectioning an object to change the perspective or viewpoint, without altering the content itself)	In a computer-based anatomy lesson, learners manipulate 3D body models—rotating, zooming, or dissecting them to explore internal structures from various angles.	A VR-based chemistry lab allows learners to manipulate molecular models by scaling, rotating, and cross-sectioning them. They can observe reactions at various levels (e.g., atomic, molecular, macroscopic), toggling between infrared, ultraviolet, and X-ray views to explore structural and energetic changes.
Level IV	**Manipulating the component content** - create non-predefined outputs by selecting and entering relevant inputs into the system	Inputting different variables into a simulation to observe how they affect the outcome.	In a physics simulation, users can adjust variables like mass, velocity, or force vectors and observe how these changes impact the behavior of virtual objects, effectively creating new scenarios within the simulation.
Level V	**Constructing the object or representations from user input -** able to generate non-predefined outputs by selecting relevant inputs and entering them into the system.	Representative examples include the use of Roblox, Minecraft, or Second Life to design and construct virtual structures and environments from scratch.	In a world-building app, users generate unique VR landscapes, buildings, and objects. These can be projected into the physical world via AR, enabling manipulation of scale and orientation. AR-VR integration merges digital design with physical interaction, which may support creativity and spatial understanding.
Level VI	**Constructing the object or contents of the representation and receiving intelligent feedback from the system through manipulative action -** Users construct objects or content within a system and receive real-time, context-aware feedback. Dynamic interaction enables adaptive exchanges between user actions and system responses.	An adaptive learning platform that analyzes student interactions, adjusts task difficulty, recommends personalized learning paths, and provides targeted instructional feedback.	A surgical training program that analyzes the trainee's technique, provides real-time corrective advice, and adjusts the difficulty of the procedure based on analysis of learner performance is an example of the adaptive and responsive nature of the AI.

 © ISLS