Chen, X.; Zhang, Y; Wu, W.; Luo, Y. & Shi, Y. (2022). Exploring assessment approaches of STEM learning in extended reality environments. *International Journal of Technology in Teaching and Learning*, *18* (2), 120-136.

Exploring Assessment Approaches of STEM Learning in Extended Reality Environments

Xiaojun Chen
St. John's University

Ying Zhang
Robert Morris University

Wei Wu
California State University, Fresno

Yupeng Luo
California State University, Fresno

Yangming Shi
University of Alabama

Abstract: This paper explores assessment approaches within the unique context of Extended Reality (XR) learning environments, which offer immersive and contextualized experiences. Focusing on STEM disciplines, the paper presents four distinctive XR-based assessment cases, exploring the discipline, context, purpose, and results of each. These insights aim to contribute to a richer understanding of learning assessment in XR environments and provide direction for future assessment designs in this emerging field.

Keywords: assessment, extended reality, virtual reality, XR learning environments, STEM education

---

Dr. Xiaojun Chen is Associate Professor in the Department of Curriculum and Instruction, School of Education, St. John's University, and can be reached at chenx@stjohns.edu. Dr. Ying Zhang is Associate Professor in the Department of Education in School of Nursing, Education and Human Studies, Robert Morris University, and can be reached at zhang@rmu.edu. Dr. Wei Wu is Associate Professor and Chair of the Department of Construction Management in Lyles College of Engineering at California State University, Fresno, and can be reached at weiwu@mail.fresnostate.edu. Dr. Yupeng Luo is Professor in the Department of Construction Management in Lyles College of Engineering at California State University, and can be reached at viluo@csufresno.edu. Dr. Yangming Shi is Assistant Professor in the Department of Civil, Construction and Environmental Engineering at the University of Alabama, and can be reached at shiyangming@ua.edu
DOI: https://doi.org/10.37120/ijttl.2022.18.2.03

INTRODUCTION

The use of extended reality (XR) has taken place in K-16 classrooms. XR is a term that generally covers a wide range of tools that enhance human technology interactions, including Virtual Reality (VR), Augmented Reality (AR) and Mixed Reality (MR), and. VR can be defined as a complex technology system that simulates or imitates real or imagined worlds (Mikropoulos & Natsis, 2011). VR allows the possibility for the users to completely shut down the outside world while immersing themselves in the virtual environment (Loomis et al., 1999). AR, on the other hand, does not isolate the users in the virtual environment but brings added virtual content to the real world. MR blends elements of VR and AR to create an environment where users can interact with real-world objects in a virtual space, and/or interact with virtual objects that can exist in the real world. Research in XR applications has shown the terms, namely, virtual reality or extended reality are often used interchangeably, but researchers have also used the different degree of virtuality to distinguish VR, AR and MR.

Di Natale, Repetto, Riva and Villani (2020) categorized three levels of virtual environment: non-immersive virtual environment, such as desktop virtual reality through 3D world on a computer; semi-immersive virtual environment, such as fulldome or smart glasses which provides a wide range of view or through wearable devices; and immersive virtual environment, such as headset and head-mounted display. XR environments provide a unique learning experience, which is significantly different from multimedia learning environments. One major difference is that XR environments provide an immersive experience, where learners can use his/her all senses to virtually go through the learning process (Di Natale et al., 2020; Bailenson, 2018). Makrnasky and Petersen (2021), for example, argue that immersive VR in particular, provides presence and agency. Di Natale and colleagues (Di Natale et al., 2020), based on their work on a systematic literature review of 18 empirical studies, also argue that immersive virtual learning environments allow for contextualized and situated learning opportunities for students in which learners can take the self control. Di Natale et al., 2020 also states that immersive virtual environments provide engaging experiences, which promote motivation and enhance learning transfer.

Studies across disciplines show that XR has significant potential to improve students' learning attitude and effectiveness (Tang et al. 2020), raise student motivation (Buchner & Zumbach, 2018; Chiang et al., 2014; Huang et al., 2010; Makransky et al., 2019; Parong & Mayer, 2018;), improve content learning (Utami & Lutfi, 2019; Wu et al., 2020), increase their identification with the STEM community (Peck et al. 2018; Starr et al. 2019). Studies have been conducted to explore and compare the impact of different VR modalities on STEM learning (Mohammadiyaghini, 2021; Shi, Du & Worthy, 2020)). A number of studies focusing on the affective aspect of learning also show that immersive virtual learning environments bring joyment, interest, and higher sense of agency (Makransky, et a., 2019; Makransky & Lilleholt, 2018; Makransky et al., 2020). Some systematic literature review also summarizes the benefits of using XR learning environments, which include the increased skill acquisition, knowledge recall, and positive emotional response (Kavanagh et al., 2017; Papanastasiou et al., 2019).

LEARNING AFFORDANCES AND ASSESSMENT OF XR LEARNING ENVIRONMENT

Recent reviews have discussed the learning affordances of virtual learning environments (VLEs). Di Nantale et al.(2020) shared four unique aspects of learning affordances of virtual learning environment as defined by Dalgarno and Lee (2010):

"3-D VLEs can be used to facilitate learning tasks that lead to the development of enhanced spatial knowledge representation of the explored domain" (p.2009),

"3-D VLEs can be used to facilitate experiential learning tasks that would be impractical or impossible to undertake in the real world" (p.2010),

"3-D VLEs can be used to facilitate learning tasks that lead to increased intrinsic motivation and engagement" (p.2010), and

"3-D VLEs can be used to facilitate learning tasks that lead to improved transfer of knowledge and skills to real situations through contextualisation of learning" (p.2011).

These affordances highlight the learning opportunities that a virtual learning environment could bring to the realm of teaching and learning, which would be guiding the designing of the learning environment as well as assessing students in learning in such environments. In terms of using XR in teaching, the principles of instructional design can be taken into consideration. Studies show that some instructional and scaffolding strategies used in less immersive teaching and learning environments may be useful in XR environments (Makransky et al., 2021). It is concluded that XR can be more effective in enhancing students' learning outcomes if certain instructional and scaffolding strategies are appropriately designed.

On the other hand, how to assess students' learning in XR environments remains a less-discussed topic. Assessment in less immersive environments generally follow the following guidelines (Brookhart & Nitko, 2019): First, assessment would need to be valid, accurate, and closely aligned with learning objectives. Second, assessment needs to be reliable. Ideally, multiple assessment methods can be utilized to ensure that the assessment provides trustworthy and accurate results. This could be achieved by including a variety of assessment formats and by conducting the assessments at different times. Third, assessment would need to provide fair and equal opportunities for students to demonstrate their learning. For example, when assessing culturally and linguistically diverse students, the linguistic difficulty levels would need to be taken into consideration to ensure these students are not punished because they cannot understand the assessment. Finally, the assessment would need to be carefully and accurately interpreted and be used to direct future instruction.

All these common principles are still applicable in designing XR-based assessment. However, due to the unique nature of XR learning environments, which emphasizes the sense of presence and agency (Makrnasky & Petersen, 2021), XR-based assessment needs to capture and reflect these unique characteristics and its alignment with learning objectives, learning affordances, as well as the overall learning contexts. For example, in XR learning environments, students may directly interact with objects and other learners. The interaction itself, which includes action, speech, and even eye gaze, may imply engagement, attention, and learning. Assessment data that capture the interaction during the learning process as a type of formative assessment thus may provide even more relevant and valuable information to the instructors or the researchers. Learning, under this unique learning environment, would be viewed more than merely as a change of memory, but a more holistic process which results changes in affective, cognitive, behavioral, and even biological domains. Recent research has explored innovative assessment methods in XR learning environments. For example, one study used virtual reality simulations to assess the competency for emergency medicine learners using virtual reality simulations (McGrath, et al., 2018). Additionally, another study involved the use of self-report measures, such as learner perception questionnaires, to evaluate the effectiveness of an augmented reality application for learning about plant anatomy (Li & Liu, 2022). These studies demonstrate the potential of different assessment methods in evaluating learning in XR environments, and emphasize the importance of utilizing a holistic approach to assessment to capture the full extent of learning outcomes.

Another point worth investigating are the types of data and assessment that could potentially be collected in the XR environment. Empirical studies in different fields have shown evidence that a variety of data could be collected in the XR environment, including biological data (Hasenbein et al., 2022), physiological data (Vesisenaho et al., 2018), self-reported data (Harron, Petrosino & Jenevein, 2019), performance data (Mead et al., 2019) and even others. Many studies adopted the mixed methods by using both qualitative and quantitative data (Han, 2020; Mead et al., 2019).

PURPOSE

In light of understanding how assessment could be designed in the XR environment, and its impact on students learning, this paper shares four cases from STEM disciplines as a platform to allow both teacher educators and STEM educators to share and discuss current practices involving XR-based assessment.

CASES

In this section, we will share four cases conducted from different disciplines piloting assessment approaches in XR learning environment. Each case is discussed in the following areas: discipline, purpose/aim, participants, data collection, data analysis and impact/results.. These four cases are: 1) tiny house VR/AR walkthrough, 2) MR-enhanced wood frame lab, 3) virtual field trips for AEC education, and 4) pipefitting in multimodality learning environment. The first three cases took place in a public university in west coast region of the USA, while the fourth case took place in a southern public university in the USA.

CASE 1: TINY HOUSE VR/AR WALK THROUGH

Discipline: *Tiny House Walk Through* case was conducted in the discipline of construction management, architectural studies, and civil engineering in higher education.

Purpose/aim: To assess the effectiveness of using VR and MR environments to develop technical skills and professional judgment in accessibility design for students in construction management, and to evaluate whether VR/MR learning environments could bridge the gap between student novices and industry professionals.

Participants: To study the assessment effectiveness of this approach, student novices and professional experts in the specified disciplines were both involved. Both groups of participants were challenged to review and assess the accessibility of a tiny house design in a simulated environment.

Description of the XR environment: This case was conducted using both virtual reality (VR) and mixed reality (MR) environments. Participants were challenged to review and assess the accessibility of a tiny house design in a simulated environment. Participants applied both technical knowledge and experience-based professional judgment in redesigning the tiny house to meet accessibility criteria. The virtual mockups were created using Unity 3D and published via the interfaces of HTC VIVE and Microsoft HoloLens, resulting in four possible virtual mockups (two for each technology). This approach enabled researchers to obtain authentic responses from participants in situ, as they interacted with the virtual environments.

Assessment Approach: In this case, participants were asked to conduct a design review and assessment for a randomly selected virtual mockup using both HTC VIVE and Microsoft® HoloLens. A graduate research assistant (GRA) provided guidance during the process, and participants provided informed consent for the activities to be recorded, including participants' movements, interactions, and comments using appropriate means. To fully utilize the unique visualization capabilities of VR and MR environments, the MR simulation experience included participants physically navigating in an actual wheelchair. This allowed participants to better evaluate the accessibility of the design. In contrast, the VR environment used a point-and-click navigation approach to simulate physical navigation, with participants seated at wheelchair height to provide a more authentic experience. The use of physical navigation in the MR environment and virtual simulation in the VR environment aimed to provide a comprehensive assessment of the virtual mockup's accessibility design. By utilizing different approaches to navigation and evaluating the design from multiple perspectives, researchers could better understand the strengths and limitations of these technologies in assessing accessibility. Figure 1 shows the participant engaging in the virtual simulation.

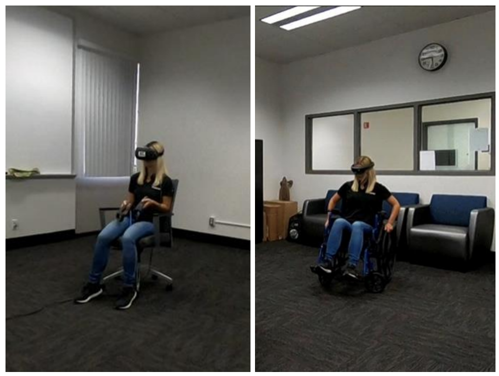

Figure 1. Participant Engaging in the Virtual Simulation of the Tiny House VR Walkthrough in Chair and Wheel Chair

Assessment Data Collection & Analysis: A variety of data were collected, including 1). verbal communication following a pre-designed think-aloud protocol focusing on the "how" and "why" aspects of the tiny house design; 2). Behavioral data to capture the key physical movements involved when evaluating the tiny house design; and 3). Pre- and post-survey to measure the participants' perceived self-efficacy in accessibility design review and assessment, tacit knowledge gains, and apprenticeship development.

The verbal communication data were captured through multiple audio (e.g., Zoom® H4nSP 4-Channel Handy Recorder) and video (e.g., Samsung® Galaxy Tab S2 and iPad 3) recording devices. The combination of audio and video recordings provided rich data for both behavioral and qualitative analysis, enabling a thorough evaluation of the effectiveness of the virtual environments and the participants' experiences within them. The behavioral data were analyzed using the Behavioral Observation Research Interactive Software (BORIS) (Friard & Gamba 2016). A behavioral coding scheme was developed to identify and evaluate the verbalized comments and key physical interactions of the participants. By assigning codes to each identified comment or interaction, a numeric value was assigned that was later used for statistical analysis in BORIS. The coding scheme was dependent upon the experience and apprenticeship relevant to the experiment, with varied codes assigned to collect different information from the experiment. The pre- and post-test surveys investigated perceptions from two different perspectives, including self-efficacy in design review and assessment, and perceptions toward affordance of VR and MR technology. Questions were designed with Likert-type scales. Specifically, for self-efficacy in design review and assessment, the Wilcoxon Signed Rank Test (Rosner et al. 2006) was used for the paired pre- and post-test data of both novices and experts. For affordance of VR and MR in design review and assessment education and application, the Mann–

Whitney U Test (MacFarland & Yates, 2016) was used to compare perceived affordance of VR and MR between novices and experts.

Results & Implications: From a behavioral analysis perspective, the performance of students in design review tasks demonstrated comparable patterns to industry professionals. However, the paths taken by them to achieve the same design assessment results were notably different. These findings suggest that faculty plays a crucial role in designing an appropriate learning approach that explicitly leads students to develop conditionalized knowledge. The assessments results from this case confirmed the educational and application affordance of VR and MR for both novices and experts, providing a solid foundation for collaborative planning of VR and MR adoption in the construction industry. Based on these findings, innovative programs leveraging VR and MR could be proposed to bridge experience-incurred gaps in design and construction education, accelerate novice expertise development, and cultivate a skilled workforce.

CASE 2: MR-ENHANCED WOOD FRAME LAB ASSESSMENT STRATEGY

Discipline: MR-enhanced wood frame lab case was conducted in the discipline of construction management in higher education.

Purpose/aim: To compare the effectiveness of traditional paper-based design communication with that of MR mockups in an outdoor wood frame construction lab, and to examine how the use of MR technology impacted student behavior and perception when participating in professional practices, and to identify factors and constraints that could affect the pedagogical use of MR in designing appropriate apprenticeship learning experiences in the undergraduate CM curriculum.

Participants: Lower-division undergraduate students majoring in construction management and civil engineering, approximately 30 divided into the morning sessions and afternoon sessions.

Description of the XR environment: The lab consisted of two parallel sessions, one in the morning and one in the afternoon, with three teams of 4 to 5 students per team in each session. The afternoon session was designated as the control group and used paper drawings for design communication, while the morning session was the experiment group and used two Microsoft® HoloLens devices loaded with the SketchUp Viewer application for MR mockups. The HoloLens devices projected the wood frame model as an interactive MR mockup, allowing students to manipulate and explore the design in various ways. They could examine specific parts of the structure using the Scene menu, review the model from different scales and perspectives using the Scale, Move, and Rotate functions, and obtain material sizes and positions directly from the model using the Measure tool. Additionally, multiple students could inspect the design at the same time using the Collaborate function. Figure 2 shows participants exploring the wood frame design during XR enhanced lab sessions.

Figure 2. Participants Engaging in XR-Enhanced Wood Frame Lab Sessions

Assessment Approach: In this case, a combination of assessments was adopted from direct observation of participants behaviors, as well as self-reported data based on XR tasks, and focus group interview on the participants. To assess and compare the performance of student teams between the control and experiment sessions, a production rate was calculated as a percentage of completion based on recorded lab progress at the end of each session.

Assessment Data Collection & Analysis: A variety of data were collected in the study, including 1). Video recordings of the student behaviors related to how they utilized the technology, and how the actual installation was influenced by the technology. This was done through a pair of GoPro® HERO6 Black video recording devices deployed at the construction yard. These cameras were positioned to record activities during both morning and afternoon sessions from opposite ends to ensure complete coverage of the installation area. Due to interference between teams and background noise from power tools and physical activities, audio recordings were not considered 2). Pre- and post-survey focusing on the students' perceptions of the wood frame lab experience varied with or without MR intervention; and 3). Semi-structured interview with open-ended questions to capture critical feedback on observed student interaction with the MR technology and its impacts on student behavior and overall lab performance.

To conduct a quantitative analysis of the video recordings, the research team coded student behaviors into four categories, calculated and summarized the corresponding time durations they spent on each category of installation activities. The coded categories included: 1). Individual use of HoloLens (experiment teams, or E-Teams) or paper drawings (control teams, or C-Team): This refers to individual students spending time on design review to obtain wood frame technical details. 2). Concurrent use of HoloLens or paper drawings: This refers to concurrent design review by two or more students. 3). Construction: This refers to the physical installation of the wood frame. 4). Gap: This refers to the non-production time of gaps between design review and installation. The derived time duration data was then used to calculate student performance in terms of productivity, and compare patterns of technology uses.

The pre- and post-survey data was compared with descriptive statistics to deduct perceived apprenticeship learning gains. The pretest established students' self-evaluated baseline understanding of plan reading, building materials, and wood frame construction before participating in this lab project. The posttest revisited the same set of questions to assess students' perceived apprenticeship learning gains via physical construction activities. A delta was calculated to identify where MR had the biggest impact on learning gains reported. The semi-structured focus group was recorded and transcribed, with a thematic analysis performed to identify key factors that contribute to the pedagogical use of MR in outdoor construction activities as part of an active learning design.

Results & Implications: The results of this study revealed that the use of MR did not offer the expected benefits in facilitating students' apprenticeship learning in outdoor physical construction activities. This finding contradicts existing research literature, but given the gaps in understanding the performance of MR in outdoor environments, it is important to identify the factors that may have influenced the pedagogical use of MR in this and future research efforts. The factors that affected students' apprenticeship learning experience with MR/HoloLens fell into three categories: technology-related, environment-related, and training-related factors.

CASE 3: VIRTUAL FIELD TRIPS FOR AEC EDUCATION

Discipline: This study was conducted with students in architecture, engineering, and construction (AEC) disciplines in higher education. VFT was used in construction field in the past, mainly through 2D and 360-degree photos. The new prototype featuring location-based learning with immersive VR technology allows learners to see the progress of an ongoing project and consists of a wide range of media types.

Purpose/aim: To evaluate the effectiveness of using the VFTs in assisting students' knowledge acquisition of field operations and construction practices.

Participants: The study involved 99 students from different courses, including upper-division courses in construction management and civil engineering and lower-division courses in construction management.

Description of the XR environment: This VFT prototype immerses students in an active construction site through a diverse array of media, including 3D models, field-captured 2D and 360-degree photos, audio and video recordings of field production and installation processes, as well as text and PDFs. Figure 3 shows the VFT displaying an active construction site scene and a 3D model scene side by side. The virtual tour provides an engaging and self-guided experience, showcasing common field practices such as excavation, concrete pouring, and steel framing. Traditionally, AEC instructors organize field trips to construction sites, covering similar topics. However, these trips can be time-consuming and require multiple visits to observe activities at different stages of construction. The proposed VFT offers a more accessible and efficient alternative, allowing students to gain a comprehensive understanding of field operations through a self-guided virtual environment. As a result, students can learn at their own pace, revisit challenging concepts, and gain valuable insights into construction practices without leaving the classroom. Overall, the VFT prototype demonstrates the potential of VR technology to transform the way AEC students learn and engage with real-world construction processes.

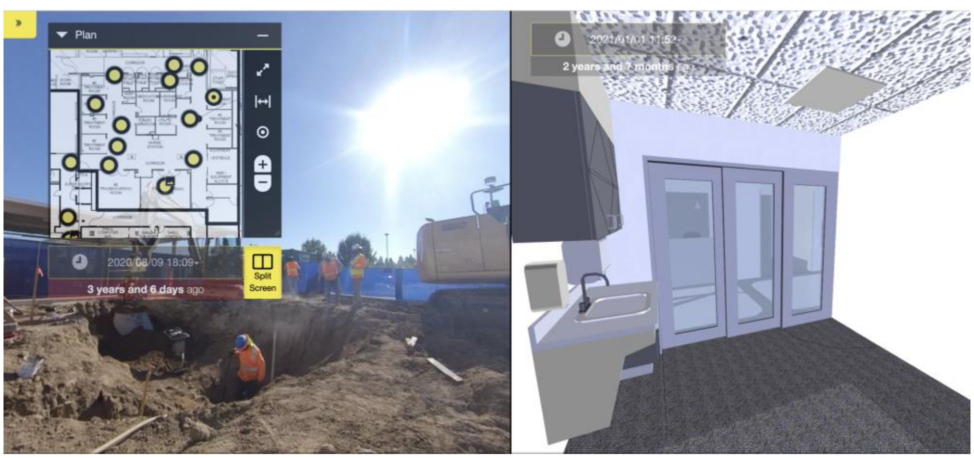

Figure 3. A SplitScreen View of the VFT Displaying an Active Construction Site Scene and a 3D Model Scene

Assessment Approach: A mixed-method approach was utilized to gather both objective and subjective data, allowing for a more comprehensive understanding of the VFT's effectiveness in AEC education.

Assessment Data Collection & Analysis: The assessment consisted of a pre-test and a post-test, with eight technical questions covering the three main topics addressed in the VFT: concrete placement, steel erection, and safety. The post-test also included four perceptional questions, aimed at eliciting feedback on the tour's highlights/takeaways, ease of navigation, overall experience, and recommendations for future improvements. A paired t-test was conducted to assess the VFT's impact on students' technical knowledge acquisition. A single-factor ANOVA F-test was used to determine if there were significant differences in the ratios of post-test grades to pre-test grades across different classes. Additionally, a cumulative frequency analysis was performed to determine the VFT's overall impact on student learning.

Results & Implications: The assessment results indicated that the proposed VFT was highly effective in facilitating technical knowledge acquisition among the students, with significant improvements observed in post-test scores. Moreover, the VFT was found to be particularly beneficial for medium-level students, contributing to their enhanced performance. These findings provide compelling evidence of the VFT's potential to improve AEC education, as well as valuable insights into its optimal implementation for different student cohorts.

CASE 4: PIPEFITTING IN MULTIMODALITY LEARNING ENVIRONMENT

Discipline: The pipefitting case was conducted in the discipline of facility management and maintenance contexts since industrial pipeline maintenance professionals especially novices often need to memorize the operational instructions before performing tasks in some confined working spaces.

Purpose/aim: To compare and assess the effectiveness of virtual learning environments in developing knowledge and motor skills in facility management and maintenance contexts compared with traditional methods, including two-dimensional isometric drawings and three-dimensional models.

Participants: A total of 90 participants, aged between 18 and 45 years participated in the study with 48 males and 42 females. According to the experimental back questionnaire, most participants had limited VR experience, and most participants had minimal prior knowledge of HVAC systems before the experiment. During the experiment, most participants experienced mild and tolerable discomfort while performing the tasks in the virtual environment, and none requested to terminate the experiment.

Description of the XR environment: The goal of this experiment is to compare the effectiveness of the visual inputs for understanding complex pipefitting instructions. Participants were asked to memorize sequences for turning or closing the valves before they replaced the plate heat exchanger in the virtual environment. There are 10 pre-start-up sequences to cut off the hot water and cold water. The virtual learning environment for 3D and VR groups was designed by using Unity 3D game engine and was displayed by the HTC VIVE HMD. There were three treatment groups: a two-dimensional (2D) isometric drawings group, a three-dimensional (3D) model group, and a VR learning group. The operation instruction of the 2D group was designed as a 2D isometric drawing of the plate heat exchanger with bulleted text operating instruction narratives on the monitor. The operation instruction of the 3D group was designed as an interactive 3D model of the plate heat exchanger with bulleted text operating instruction narratives. The participants in the 3D group could use the keyboard and mouse to change the view of operating instruction texts and the 3D model. The operation instruction of the VR group was designed to use a Head Mounted Display (HMD) headset to review the operating instructions as well as a virtual plate heat exchanger model in an immersive virtual environment. Participants in the VR group could also interact with the plate heat exchanger model while reviewing the operating instructions. Figure 4 shows the participant performing the pipefitting tasks in the virtual environment.

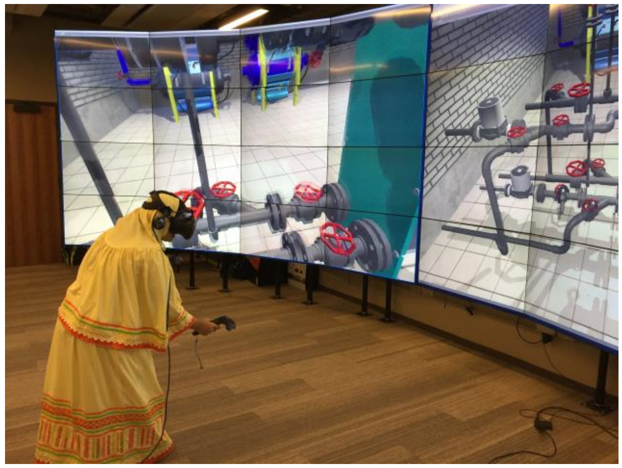

Figure 4. View of Participant Performing Pipefitting Tasks

Assessment Approach: To achieve the assessment goal, a between-participant design was adopted with three treatments. The participants were randomly assigned to one of the three groups before the experiment. The research investigators ensured that participants' eye movements were accurately captured by the eye trackers that integrated with the monitor and the HMD after several five-point calibration trials. Participants were also given instructions about how to use the two controllers to interact with the virtual objectives in the virtual environment. There were two main sessions in the experiment, which were the review session and the operation session. In the review session, the participants were instructed to review and memorize the pipe maintenance sequence in 5 minutes. In the operation session, the participants were asked to perform the pipe maintenance task in the VR environment without any time limit. At the end of all experimental stages, participants were asked to fill out a post-questionnaire to provide comments and feedback.

Assessment Data Collection & Analysis: In this case, the participants' learning performance was assessed from different dimensions, including learning performance and eye-tracking data. For the learning performance, we evaluated it from two perspectives: accuracy and time. The accuracy was defined as the accuracy in performing correct steps compared to total steps and directly represents how well the participants memorized and performed the task. Time was defined as the time that participants used to complete the task. The time represents how efficiently the participants finished the task. For the assessment from the eye-tracking data, we extracted three features: review fixation ratio, gaze entropy, and pupil dilation related to learning performance. The learning performance data were analyzed by statistical methods. we used a non-parametric test (Kruskal-Wallis) to compare learning performance across different groups. For the eye-tracking data, we used a two-way non-parametric test (Scheirer-Ray-Hare) to compare the review fixation ratio across three groups. For the gaze entropy, we found that participants in both the 3D group and the VR group had higher gaze entropy than the 2D group. Since the gaze entropy represents the randomness of the gaze movement between different AOIs, these results may indicate that participants in the 3D group and VR group had more frequent gaze transitions between graphical information and text information. It may lead to a better development of learning performance. For the results of pupil dilation, participants in both the 3D group and the VR group had significantly higher pupil dilation than the 2D group. During the analysis of the eye-tracking data and learning performance data, we also found a positive correlation between the eye-tracking features and the learning performance. The results indicated that more frequent gaze entropy (movement) has led to better development of learning performance during the review phase.

Results & Implications: The results suggested that the 3D and VR groups performed better than the 2D group in operation accuracy, but participants had similar operation times across the three groups. These results revealed that participants in the 3D and VR groups might have better learning performance than the 2D group. It was also found that participants in the 3D group and VR group had more frequent gaze transitions between graphical information and text information. It may lead to better development of learning performance.

DISCUSSION

XR learning environment is unique as it provides immersive, contextualized, and situated learning experiences to learners (Di Natale et al., 2020; Bailenson, 2018; Kavanagh et al., 2017; Papanastasiou et al., 2019). Learners can immerse themselves in simulated environments, practice their skills in realistic contexts, receive real-time feedback and guidance, and take control of their own learning processes. These environments can enhance learners' engagement, motivation, and learning outcomes, making them a valuable tool for education and training. The XR learning environment creates a potential for learners to gain experiences which would be difficult or impossible to undertake in the real world. It emphasizes a sense of agency (Makrnasky & Petersen, 2021), where learners can take control of their own learning processes. These unique affordances of XR learning environments require a sophisticated design of assessment to obtain accurate information needed. We propose that when designing assessment in XR learning environments, the instructor/researcher needs to align the assessment with learning goals and incorporate a variety of data types.

ALIGNING ASSESSMENT WITH LEARNING GOALS

The first key point in designing assessment in XR learning environments is that the data points would need to be carefully selected to closely align with learning goals and/or objectives. This general principle in assessment design not only works in traditional learning contexts, but applies in XR-based learning environments. The learning goals and objectives determine the assessment methods and data points, not the vice versa. Well-designed assessments are valid and reliable, which provides essential feedback to further guide the learning process. When designing assessment in XR learning environments, a clear understanding of how assessment methods and data points can provide valuable information to the instructor/researcher would need to be identified.

Figure 5 below shows a visual representation of how assessment may be designed in XR environments. The center of the assessment design is the learning goals and/or objectives. Although we propose that a variety of data points would need to be incorporated into the assessment design, we advocate a careful selection of data types. Each data point would need to be directly related to the learning goals and/or objectives.

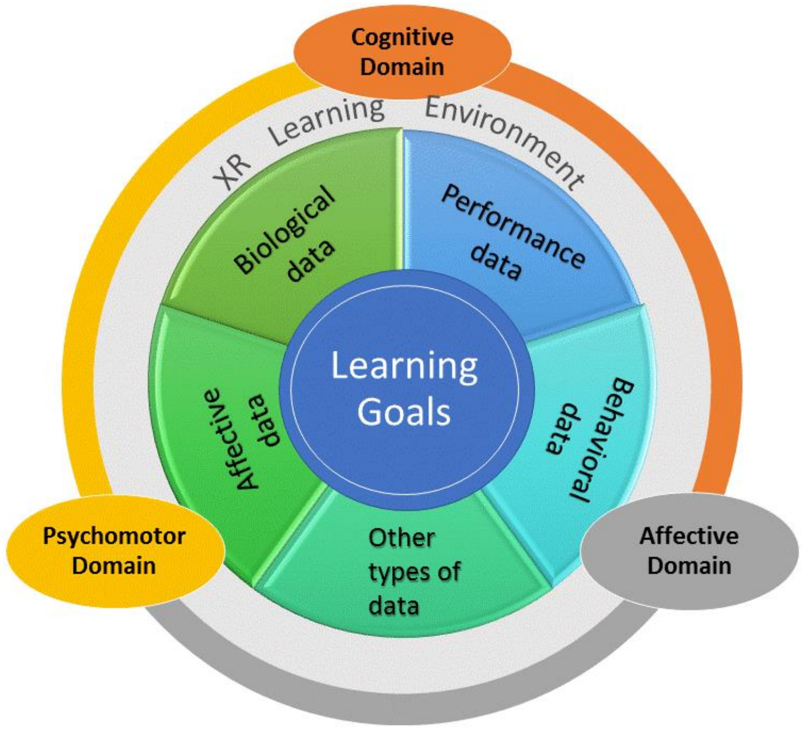

Figure 5. Visual Representation on Assessment Design in XR Environments

Once the learning goals and objectives are clearly defined, a variety of data can be collected, including the cognitive data, affective data, behavioral data, biological data, and even other types of data. Different data collection methods, such as survey, questionnaire, interview, pre- and post-tests can be used, depending on the learning goals and objectives. The outer circle of the visual representation shows three learning domains, cognitive learning, affective learning, and psychomotor learning. It needs to be pointed out that different data may address one or multiple learning domains. For example, biological data, which may include eye-tracking data as described in case 4, can be analyzed to address affective learning and cognitive learning.

One caution we keep in mind when designing the visual representation is that besides the cognitive data, affective data, behavioral data, and biological data found in the literature and cases described above, there might be other types of data that can be used to address different learning domains as well. For example, when studying how students learn foreign languages in XR environments, linguistic data may be added to address their cognitive learning (Lee, 2013; Chen, 2016).

INCOPORATING MULTILAYER OF DATA TYPES

The second key point in designing assessment in XR environments is the use of multiple assessment methods and data points to capture the different aspects of learning. For example, in addition to traditional measures of knowledge acquisition, such as multiple-choice questions or written responses, assessments in XR environments may also include behavioral observations, performance tasks, and affective data, such as learner perception questionnaires.

This approach recognizes that learning in XR environments is complex and multifaceted and cannot be evaluated through a single method or data source. One major difference between learning in non-XR environments vs. XR environments is that the XR learning environments provide more immersive, contextualized, and personalized learning in which the learners take a strong control on their learning process (Di Natale et al., 2020; Makransky & Petersen, 2021). Because the learners' sense of agency and immersive experience can strongly influence their learning outcomes, data points that capture the affective and psychomotor aspects of learning should be included when assessing the overall learning effectiveness in XR learning environments. For example, learner perception questionnaires can capture learners' attitudes and perceptions toward the learning experience, which can provide insight into the effectiveness of the learning design in XR environments. Additionally, biological data, such as heart rate variability or skin conductance, can provide objective measures of learners' emotional states during the learning experience.

By utilizing a variety of assessment methods and data points, designers of XR learning environments can gain a more comprehensive and holistic understanding of the learning in these environments. This approach allows for a more accurate and comprehensive evaluation of the learning outcomes, which can inform future design decisions and ultimately lead to better learning experiences for learners.

CHALLENGES AND OPPORTUNITIES

Although XR-based learning shows potential in motivating students for more contextualized and individualized learning (Di Natale et al., 2020; Bailenson, 2018; Kavanagh et al., 2017; Papanastasiou et al., 2019), it has not been widely applied across different contexts. First of all, XR-based assessment is costly. The assessment often involves expensive software and hardware, which limits the possibilities for instructors or researchers. Using the XR technology also requires a significant amount of training, which takes a lot of time. Research shows teachers, for example, often find it difficult to embrace XR technology due to a variety of factors, such as lack of administrative support and physical discomfort of using the technology (Bahng & Lee, 2017; Swier & Peterson, 2018; Vesisenaho et al., 2018).

Besides these obvious challenges in XR learning, there are some unique challenges identified in designing XR-based assessment. One challenge is how to design appropriate assessments to capture the learning process within the XR environment. Traditional data points, such as pre- and post-tests capture the end product of learning, but not the process of learning within the XR context. To understand how learners make sense of the learning in the XR environment, multiple assessments embedded in the environment need to be designed to capture the learning progress. One possible solution to this issue is to include multiple data sets, especially data collected while students are actively engaged in learning in the XR environment. As shown in the cases described above, it is recommended that data sets such as observational data or eye-tracking data can be incorporated, as they provide valuable information about the learning process.

On the other hand, due to the large amount of data (in different types and forms) needed to closely reveal the learning progress, the data analysis process needs to be considered when designing the XR-based assessment. One challenge is how to select the data so the amount remains feasible for analysis. For example, when collecting behavioral data, video recording is often used. However, it is both difficult and unnecessary to analyze hours of hours of recording. Criteria to select the data points based on the learning goals and/or objectives need to be identified to ensure the data can be analyzed in a purposeful manner.

Having a clear criterion in data selection also ensures that when reporting the findings, supporting evidence can be pulled across multiple data sets.

CONCLUSION AND FUTURE RESEARCH

Teaching and learning in XR environments become more and more popular across different disciplines in K-16 education, especially in science, engineering, medical, and other fields. XR environment provides immersive, engaging learning experiences, which is difficult or impossible in real classroom settings (Dalgarno & Lee, 2010; Bailenson, 2018). The unique affordances of XR learning environments pose challenges in assessment design.

General assessment principles, such as assessment would need to be closely aligned with learning objectives, assessment would need to be valid and reliable, and assessment would need to provide equal opportunities for learners to demonstrate their learning outcomes, are still applicable to assessment design in XR environments. However, as we have proposed above, several key components need to be carefully addressed in assessing learning in XR environments. These include the use of multiple data sets to as accurately as possible capture the learning processes and learning outcomes, and the selection of each data type to closely align with the learning goals and objectives. The assessment design in XR environments may encounter some challenges as well. For example, leaning how to use XR devices and tools can take a significant amount of time. In addition, the data analysis may be more complex as multiple data sets are used.

In conclusion, designing assessments in XR environments require a clear understanding of the learning goals and objectives, careful selection of data types to capture the learning outcomes and learning processes, and sound data analysis methods to pull different data points to answer the research questions. With our discussion, it’s our hope that this discussion may shed some light on future assessment design practices in examining the learning in XR environments.

In the burgeoning field of XR environments across STEM disciplines, future research could focus on prioritizing the development of assessment frameworks that reflect learning theories and integrate current assessment practices with the unique aspects of XR. Understanding the learning curve associated with XR tools, exploring different data types for varied learning outcomes, and ensuring equity and accessibility in XR assessments are vital. Ethical considerations around data collection and analysis, long-term studies on XR's effectiveness on teaching and learning, and seamless integration with traditional methods need to be addressed. Collaboration across disciplines, insights into global perspectives, and a focus on the role and professional development of educators will further contribute to a comprehensive understanding of XR's potential. These directions are integral to cultivating XR as an innovative and inclusive educational tool that aligns with the specific needs and opportunities within STEM education.

REFERENCES

Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do*. WW Norton & Company.

Bahng, E., & Lee, M. (2017). Learning experiences and practices of elementary teacher candidates on the use of emerging technology: A grounded theory approach. *International Electronic Journal of Elementary Education*, *10*(2), 225–241. http://10.0.104.198/iejee.2017236118

Brookhart, S. M., & Nitko, A. J. (2019). *Educational assessment of students*. London, UK: Pearson.

Buchner, J., & Zumbach, J. (2018). Promoting intrinsic motivation with a mobile augmented reality learning environment. In I. A. Sanchez & P. Isaias (Eds.), *Proceedings of the 14th International Conference Mobile Learning 2018* (pp. 55–61). Iadis.

Chen, Y.-L. (2016). The ffects of Virtual Reality learning environment on student cognitive and linguistic development. *Asia-Pacific Education Researcher (Springer Science & Business Media B.V.), 25*(4), 637–646. http://10.0.3.239/s40299-016-0293-2

Chiang, T. H. C., Yang, S. J. H., & Hwang, G.-J. (2014). An Augmented Reality-based mobile learning system to improve students' learning achievements and motivations in natural science inquiry activities. *Journal of Educational Technology & Society*, *17*(4), 352–365.

Dalgarno, B., & Lee, M. J. (2010). What are the learning affordances of 3-D virtual environments?. *British journal of educational technology*, *41*(1), 10-32.

Di Natale, A. F., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology, 51*(6), 2006-2033.

Friard, O., and Gamba, M. (2016). "BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations." *Methods in Ecology and Evolution, 7*(11), 1325-1330.

Han, I. (2020). Immersive virtual field trips in education: A mixed-methods study on elementary students' presence and perceived learning. *British Journal of Educational Techno- logy*, *51*(2), 420–435. http://10.0.4.87/bjet.12842

Harron, J. R., Petrosino, A. J., & Jenevein, S. (2019). Using virtual reality to augment museum-based field trips in a preservice elementary science methods course. *Contemporary Issues in Technology & Teacher Education, 19*(4), 687–707.

Hasenbein, L., Stark, P., Trautwein, U., Queiroz, A. C. M., Bailenson, J., Hahn, J.-U., & Göllner, R. (2022). Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom. *Computers in Human Behavior, 133*, 107282. https://doi.org/https://doi.org/10.1016/j.chb.2022.107282

Huang, H.-M., Rauch, U., & Liaw, S.-S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education*, *55*(3), 1171–1182.

Kavanagh, S., Luxton-Reilly, A., Wuensche, B., & Plimmer, B. (2017). A systematic review of Virtual Reality in education. *Themes in Science and Technology Education*, *10*(2), 85–119.

Li, M., & Liu, L. (2022). Students' perceptions of augmented reality integrated into a mobile learning environment. *Library Hi Tech*.

Lee, S. (2013). Can speaking activities of residents in a virtual world make difference to their self-expression? *Journal of Educational Technology & Society*, *16*(1), 254–262.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior research methods, instruments, & computers*, *31*(4), 557-564.

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, *35*(6), 691-707.

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*, 1-22.

Makransky, G., Petersen, G. B., & Klingenberg, S. (2020). Can an immersive virtual reality simulation increase students' interest and career aspirations in science?. *British Journal of Educational Technology*, *51*(6), 2079-2097.

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, *113*(4), 719.

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, *66*(5), 1141-1164.

MacFarland, T. W., & Yates, J. M. (2016). *Introduction to nonparametric statistics for the biological sciences using R* (pp. 103-132). Cham: Springer.

McGrath, J. L., Taekman, J. M., Dev, P., Danforth, D. R., Mohan, D., Kman, N., ... & Won, K. (2018). Using virtual reality simulation environments to assess competence for emergency medicine learners. *Academic Emergency Medicine*, *25*(2), 186-195.

Mead, C., Buxner, S., Bruce, G., Taylor, W., Semken, S., & Anbar, A. D. (2019). Immersive, interactive virtual field trips promote science learning. *Journal of Geoscience Education*, *67*(2), 131–142.https://doi.org/https://doi.org/10.1 080/108999 95.2019.1565285

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & education*, *56*(3), 769-780.

Mohammadiyaghini, E. (2021). *Evaluating the application of mmersive virtual reality in construction education and inspection* [Master Thesis, California State University, Fresno]. ScholarWorks.

Papanastasiou, G., Drigas, A., Skianis, C., Lytras, M., & Papanastasiou, E. (2019). Virtual and augmented reality effects on K-12, higher and tertiary education students' twenty-first century skills. *Virtual Reality*, *23*(4), 425–436. https://doi.org/10.1007/s10055-018-0363-2

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, *110* (6), 785–797. https://doi.org/10.1037/edu0000241

Peck, T. C., Doan, M., Bourne, K. A., & Good, J. J. (2018). The effect of gender body-swap illusions on working memory and stereotype threat. *IEEE transactions on visualization and computer graphics*, *24*(4), 1604-1612.

Rosner, B., Glynn, R. J., & Lee, M. L. T. (2006). The Wilcoxon signed rank test for paired comparisons of clustered data. *Biometrics*, *62*(1), 185-192.

Shi, Y., Du, J., & Worthy, D. A. (2020). The impact of engineering information formats on learning and execution of construction operations: A virtual reality pipe maintenance experiment. *Automation in Construction*, 119, 103367.

Starr, C. R., Anderson, B. R. & Green, K. A. 2019. "I'm a Computer Scientist!": Virtual Reality Experience Influences Stereotype Threat and STEM Motivation Among Undergraduate Women. *Journal of Science Education and Technology, 28*, 493-507.

Swier, R., & Peterson, M. (2018). 3D digital games, virtual worlds, and language learning in higher education: Continuing challenges in Japan. *JALT CALL Journal*, *14*(3), 225–238.

Tang, Y. M., Au, K. M., Lau, H. C., Ho, G. T., & Wu, C. H. (2020). Evaluating the effectiveness of learning design with mixed reality (MR) in higher education. *Virtual Reality*, *24*(4), 797-807.

Utami, I. W. P., & Lutfi, I. (2019). Effectivity of Augmented Reality as media for history learning. *International Journal of Emerging Technologies in Learning*, *14*(16), 83–96. http://10.0.15.151/ijet.v14i16.10663

Vesisenaho, M., Juntunen, M., Häkkinen, P., Pöysä-Tarhonen, J., Miakush, I., Fagerlund, J., & Parviainen, T. (2018). Virtual Reality in education: Focus on the role of emotions and physiological reactivity. *Journal of Virtual Worlds Research*, *12*(1), 1–15.

Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, *51*(6), 1991–2005.