Utilizing Extended Reality Usability Heuristics to Drive Effective XR Training Applications

Jessyca Derby, Claire Hughes, JoAnn Archer
Design Interactive, Inc.
Orlando, FL
jessyca.derby@designinteractive.net, claire.hughes@designinteractive.net, joann.archer@designinteractive.net

ABSTRACT

Extended reality (XR) has the potential to be a differentiator for the U.S. Department of Defense (DoD) in terms of training. XR encourages learners to be engaged in virtual content, embody tasks, and be involved in active learning by "doing." With the development and implementation of flight and battle training simulations as well as the establishment of XR cross-functional teams for the pursuit of XR training frameworks, the Army, Navy, Air Force, Marine Corps, and Space Force have demonstrated their support for this innovative technology. To make training with XR technologies adoptable, the user interface must have strong usability; it must be effective, efficient, and satisfying for the user. If usability is not considered during design, the training could become difficult to understand, frustrate the user or take more time than compared to traditional training methods, which would nullify the benefits. This paper details the results from a usability assessment of three different XR applications across multiple devices and defines improvements for enhancing the effectiveness of XR training solutions as well as provides recommendations for the successful use of heuristic evaluation methods. The authors conducted heuristic evaluations using a validated AR/MR heuristic checklist to identify strengths and shortcomings in a variety of XR training applications, including Air Force communications and maintenance operations and DoD and Department of Homeland Security Tactical Combat Casualty Care. The applications were evaluated on both tablet and head worn display. The heuristics used for the evaluations, "The Derby Dozen: 12 Usability Heuristics for AR and MR," resulted in actionable redesign recommendations that will improve the usability and learning efficacy of these applications. This user-centered design approach can aid developers of XR solutions in producing more effective and engaging training solutions to drive improved learning outcomes and increase readiness across the US military.

ABOUT THE AUTHORS

Jessyca Derby is a Research Associate at Design Interactive, Inc. She has 6 years of experience conducting Human Factors research, with a specialization in user experience (UX) assessment for extended Reality (XR) products. She her expertise in human-centered design, she creates and evaluates XR training applications that enhance usability and provide users with positive and meaningful experiences.

Claire Hughes is a Senior Research Associate and the Human-Systems Integration Portfolio Manager at Design Interactive, Inc. She has 10+ years of Human Factors Engineering experience focusing on project management, end-user evaluations, and research analysis for training and operational systems. She excels in analyzing learning needs, designing intuitive interfaces, and applying evidence-based principles to optimize the learning process. Her skills in usability testing and user experience research enable her to continuously improve educational programs based on learner feedback and performance data.

JoAnn Archer is the eXtended Reality (XR) Solutions Portfolio Manager at Design Interactive with 15 years of experience in system engineering and program management. Through leading the successful completion of numerous DOD, FAA, and commercially funded projects, JoAnn ensures XR products are developed utilizing applied research and data analytics that keep AR/VR/MR products on the bleeding edge and provide a competitive advantage when transitioning to end users. Her applied visionary approach coupled with sound system engineering ensures that the design, development, testing, and transition of XR immersive training, operational job aid solutions, and decision-making tools are optimized for their specific users, tasks, and context of use.

Utilizing Extended Reality Usability Heuristics to Drive Effective XR Training Applications

Jessyca Derby, Claire Hughes, JoAnn Archer
Design Interactive, Inc.
Orlando, FL
jessyca.derby@designinteractive.net, claire.hughes@designinteractive.net, joann.archer@designinteractive.net

INTRODUCTION

Technologies such as virtual reality (VR), augmented reality (AR), mixed reality (MR) or the mixture of all three under extended reality (XR) have the potential to be differentiators for integrated training readiness to assure deterrence from near-peer adversaries. With the development and implementation of flight and battle training simulations as well as the establishment of XR cross-functional teams for the pursuit of XR training frameworks, the Air Force, Space Force, Navy, Army, and Marines have demonstrated their support for these innovative technologies. XR integrates virtual 3D assets within a real or virtual operational space facilitating multimodal interactions with virtual and real equipment, assets, people, cues, conditions, and instructional components that build critical decision-making situational awareness and psychomotor skills. The value of XR technology to training is derived from the potential to provide authentic, appropriately realistic learning experiences that increase skill acquisition and knowledge retention. To ensure XR is effectively developed and implemented it is essential that products are deployed using AR-based instructional support for learners across the spectrum to increase cognitive readiness, VR environments for immersion to increase contextual readiness, and MR real-world interactions for internalization (muscle memory) to increase psychomotor readiness (Stanney et al., 2023). Additionally, it is essential to consider task-technology fit when specifying XR training programs, as no XR form factor or approach is "best" across all contexts.

As the Department of Defense (DoD) integrates XR into its processes, it has achieved significant improvements, including cost savings and up to a 60% increase in training efficiency (Hopkins, 2021). By continuing to leverage XR technologies, the DoD can transition from traditional training to high-velocity learning environments, benefitting from reduced training time, costs, and materials. The U.S. military is exploring various XR applications across different services, including tactical, flight, maintenance, medical, and warfighting training (Congressional Research Service, 2022). Notably, the Army's top acquisition official has approved the next phase of development for the service's "do-it-all" device, the Integrated Visual Augmentation System (South, 2023). Additionally, the Air Force Special Operations Command (AFSOC) has established an XR cross-functional team to create an XR Training Framework for the entire command (Rasmussen, 2023), and U.S. Space Force Unit 13th Delta Operations Squadron (13 DOS) will use the XR suite SOLAR to teach space operational concepts and situational awareness in the space domain (Auganix, 2023). As the DoD expands its research and development efforts in XR technology and moves these technologies through the transition phase, it is imperative to optimize the development of these technologies to maximize the benefits of these advancements.

For XR training technologies to be adoptable across the services and the enterprise, their interfaces must have strong user experiences (UX) that enhance the usability of the technology; they must be effective in driving proficiency, efficient when engaging users, and satisfying for the user. If usability is not considered during the design and development phase, the training could become difficult to understand, frustrate the user or take more time than compared to traditional training methods, which would nullify the benefits. However, traditional design and usability expectations do not directly translate from two-dimensional training platforms to the spatialized applications afforded by XR. Thus, usability studies with end users must be iteratively conducted to ensure that the XR systems align with users' mental models and provide them with an appropriate level of support and guidance. Heuristic evaluations are a usability inspection method used to proactively identify usability issues in a user interface (UI) early in the design and development cycle. These evaluations involve a small group of evaluators systematically reviewing the interface and judging its compliance with recognized usability principles (i.e., the "heuristics"; Nielsen, 1994). These evaluation frameworks are created based on usability rules of thumb and can be used iteratively as part of the design process.

However, conducting systematic heuristic evaluations early in the process is critical to the successful development of complex and innovative systems such as XR applications.

Purpose of the Current Study

The aim of this paper is to detail the results of a set of heuristic evaluations of three different XR applications across multiple devices and to assess usability and define improvements for enhancing the effectiveness of these XR training solutions. A summary of usability issues found across the different XR products and devices is included as well as recommendations for conducting heuristic evaluations to create a list of lessons learned for developing XR products.

METHOD

To systematically discover what usability issues occurred in the three XR applications, the authors assimilated teams to conduct heuristic evaluations. The results from each application are described below as their own case study. The heuristic evaluation methodology was chosen because it can be implemented during the iterative design process (these applications continue to be iterated upon), it can provide more direct feedback than users may describe during usability testing, it is faster and more cost-effective than usability testing, and it can capture similar usability problems that could be found during usability testing (Nielsen, 1994). By conducting these heuristic evaluations, the authors were able to gather and analyze usability strengths and areas of need improvement for the evaluated XR applications. These strengths and weaknesses were captured and translated into design recommendations for future XR applications.

During this study, teams of three researchers at Design Interactive (DI) conducted heuristic evaluations on each of the three case studies: Department of Homeland Security (DHS) Tactical Combat Casualty Care (TCCC) training, Air Force Communications and Maintenance Operations training, and XRMentor® (see Table 1). These applications were chosen because they provide a diverse range of use cases and form factors, including both mobile and head worn display (HWD). The aim was to evaluate the usability strengths and weaknesses across different application types and form factors to determine in which situations a user may benefit from these differences. The evaluators did not have any direct ties with product development. This was done to avoid any bias during the heuristic evaluations, as those who developed the application may rate it more positively than those who did not. The evaluators did have previous experience using the devices and/or applications in the past and knew the goals of the applications to effectively complete tasks and comprehensively complete the heuristic evaluations. One evaluator completed multiple heuristic evaluations. A total of eight evaluators were used.

Table 1. Heuristic Evaluation Case Study Summary

Case Study	Evaluator
1. DHS TCCC training (HWD – HoloLens 2)	Evaluator A, Evaluator B, Evaluator C
2. Air Force Communications and Maintenance Operations training (Mobile – iPad)	Evaluator D, Evaluator E, Evaluator F
3. XRMentor® (HWD – Magic Leap 2)	Evaluator A, Evaluator G, Evaluator H

Each evaluator was instructed to use their assigned application. All applications contained lessons that aimed to teach the user how to complete a task, so the lessons followed a linear path for the evaluator to follow. After each evaluator completed the lessons, they were given the "The Derby Dozen: an AR/MR Usability Heuristic Checklist" excel template to complete their evaluations (Augmented Reality for Enterprise Alliance, 2023; Derby, 2023). This heuristic checklist was chosen because it has been empirically validated and identifies a comprehensive list of usability issues that could occur with a variety of AR/MR applications and form factors (including both HWD and mobile devices). It encompasses digital and physical aspects of the UX that could impact the overall usability of an XR application or device such as: how easy it is to set up and maintain use of the application and/or device; if user interactions are intuitive or confusing; how well virtual holograms are integrated with the real world; how comfortable it is to use the device; and how a user's and/or bystander's privacy is protected by the application. The Derby Dozen heuristic checklist contains a total of 12 heuristics and 109 checklist items. The heuristics and a representative checklist item from each heuristic is shown in Table 2.

Table 2. Heuristics from the Derby Dozen: an AR/MR Usability Heuristic Checklist

Heuristic	Representative Checklist Item (out of the total 109 items)
Heuristic 1: Unboxing & Setting Up	When the user interacts with the device for the first time, are they introduced to the user interface, basic interaction methods, and basic features/content?
Heuristic 2: Instructions	Do instructions provide actionable feedback?
Heuristic 3: Organization & Simplification	Does the screen space focus on the virtual elements rather than controls or other non-AR/MR features, as appropriate?
Heuristic 4: Consistency	Do virtual elements act as the user would expect them to in the real world?
Heuristic 5: Integration of Physical & Virtual Worlds	Do the virtual elements help the user accomplish the required tasks in a meaningful way?
Heuristic 6: User Interaction	Does the device and/or application accommodate for the user to complete other necessary real-world tasks?
Heuristic 7: Comfort	Can the user experience the device and/or application without pain, discomfort, nausea, disorientation, etc. DURING use?
Heuristic 8: Feedback to the User	Does the device and/or application provide feedback for user input?
Heuristic 9: Intuitiveness of Virtual Elements	Are virtual elements and controls placed near objects they reference?
Heuristic 10: Collaboration	If users are sharing the same virtual space, are virtual landmarks included to help orient users who may be in different physical spaces?
Heuristic 11: Privacy	Is it clear how user data is collected, stored, used, and protected?
Heuristic 12: Device Maintainability	Are device parts fixable and replaceable as needed?

Evaluators completed the checklist by giving each item a qualitative score based on how well the heuristic was satisfied by the application/device ("Yes" the heuristic was satisfied, "Somewhat" satisfied, "No" not satisfied, or "Not Applicable"). For example, for the checklist item, "Are virtual elements that are outside of the field of view (FOV) easy to find?", an evaluator may answer "Yes" if it was very easy, "Somewhat" if it was only easy in some situations, "No" if it was never easy, and "Not Applicable" if there is never an instance where a virtual element is outside their FOV. Additionally, evaluators provided qualitative comments to describe additional information about why they rated the heuristic checklist item the way they did and included ideas they had to improve the application. Evaluators were invited to go back to the lesson as they completed the heuristic evaluation, but this was not required. Each evaluation took approximately 1 hour to complete. After the three heuristic evaluations for each application were complete, the results were compiled by the researchers, reviewed, and common themes were identified.

RESULTS

Case Study 1: DHS TCCC Training (HWD – HoloLens 2)

Information About Application

The DHS Security TCCC Training application (see Figure 1) was deployed on the HoloLens 2, an HWD. The purpose of this application is to train members of the DHS in TCCC while on assignment. This application provided the user with a variety of stressful scenarios where they had to provide casualty care. By deploying this application on an XR HWD, users can be hands-free when practicing recently learned procedural skills while also embodying situation awareness in a virtual environment.

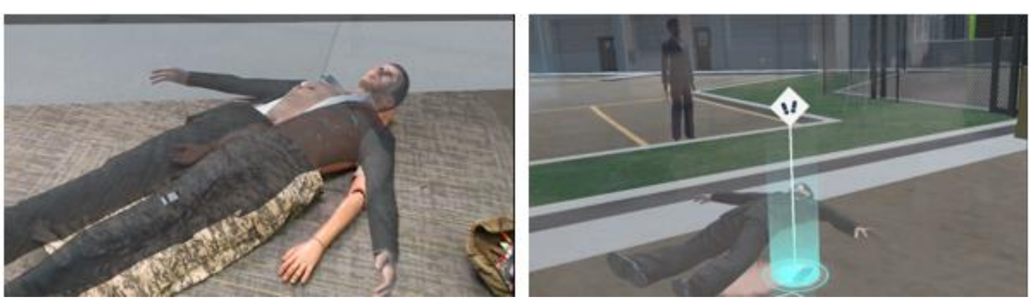

Figure 1. DHS TCCC Training Application

Takeaways from the Heuristic Evaluations

Strength 1: Keeps the User Focused by Increasing Immersion

Since this application is deployed on an AR device, the user's environment is not fully occluded like VR applications. The heuristic evaluations showed that it was still very immersive for users. This was due to the utilization of sounds, natural narrative, and a focus on XR content rather than text. When the user enters the environment, they not only see the virtual content, but they hear it as well (e.g., crowd chatter, splashes of liquid, wounded casualties, gun shots, footsteps, etc.). Additionally, characters within the scene instruct the user, asking them to complete actions directly with terms such as, "Hey, you! Go help treat that guy!" As the user progresses through the lesson, they are directed to interact with the environment by walking to different areas or providing treatment on a medical manikin. As Evaluator A pointed out, this helps the user focus on the environment, encouraging immersion, "I was very focused on the casualty in both situations rather than the menu… includes great visuals and audio to pair with it. It gets the user active by walking to navigation points as well." Developers who want their XR applications to become more immersive should build in ways to encourage users to interact with the environment as they would in the real world by instructing the user to interact with the environment directly (e.g., walk around, touch an object to pick it up, etc.) and by adding realistic environmental cues (e.g., environmental audio and realistic instructions given by avatars).

Strength 2: Avoids Overwhelming the User by Decreasing Cognitive Load

Even though the evaluators felt very immersed in a simulated environment, they did not feel overwhelmed by the application. The evaluators stated that this was because the application lessened the user's cognitive load by using familiar interactions, linear narratives, and provided brief and direct instructions when necessary. This application did not require any novel user interactions; users were only required to interact by performing native HoloLens gestures (air tap and touch) or walking around their environment. This allowed users to focus on learning how the application functioned rather than learning new controls as well. To encourage users to learn the application without feeling overwhelmed, they were introduced to their environment upon loading into the virtual scene before being asked to immediately care for a casualty. For example, they were given a scenario brief that included why they were in this environment, encouraged to look around, and were asked to walk to an area when they were ready to continue. This allowed users to gather their bearings before being thrown into a stressful situation that required them to provide medical assistance. Once the scenario began, the user was given clear goals. As Evaluator B stated, "Even without a tutorial, it is clear what the user needs to do in the application." If the user had a delay in their response, the application would give further direction such as, "What are you doing? Get to the bathroom, that is where they took your casualty." This provides more context for users who may be stuck or confused. Developers who want to decrease the cognitive load of their applications should use familiar interactions, ease users into the environment upon load in, and provide users with clear and direct goals.

Strength 3: Interacting with the Real World While Also Being Immersed in the Virtual World

Even though the application encouraged immersion, the HWD used for this application still enabled the user to be aware of their real environment. HoloLens 2 projects virtual onto a clear glass lens situated in the user's FOV, which creates translucent holograms that still allow the user to see through to the real environment. This did not negatively impact the amount of immersion evaluators felt, since audio and visual cues still provided a sense of "being there," and this made it easier to interact with the real environment. Evaluators included the following comments that describe this, "I was able to easily walk around physical objects that were in my way" and "It is somewhat translucent which makes it easy to perform the task. I can also easily see under the HoloLens, can flip it up, or turn down brightness if it becomes an issue." This balance is critical for training when users need to interact with their real environment (such as applying treatment interventions on a medical manikin) and for safety if they are walking around obstacles. When creating XR applications, developers should use optical see-through or video pass-through technology to enable the user to view the real environment around them while still being immersed in the virtual world.

Area of Improvement 1: A Need to Direct the User's Field of View Towards Their Goals

Instructions to the user should not only be understandable, but should also be congruent with the visuals presented. Evaluator B pointed out that this was not always the case when interacting with this training application, "The instruction said, 'keep a close eye on your target,' and I saw him on the screen, so I walked closer to him. When nothing happened, I was confused. After looking around, I figured out that I needed to go to a navigation point that was not on the screen (I had to look down to see the path to it) which was confusing." The user's location and what is visible in their FOV should be considered when creating the instructions and narrative. Since the evaluator's target was visible immediately, but the navigation point was not, when the target was referenced, the evaluator anchored themselves on what they could see, and walked to the wrong location. To avoid this, developers should reference what the user is likely to see in their immediate FOV or provide navigational cues that direct their attention to the correct location.

Area of Improvement 2: A Need to Set User Expectations Early

A strength of this application was that it eased the users into the virtual environment and kept them immersed while in the scenario. However, some evaluators still stated that they felt taken aback when the environment started, as Evaluator C noted, “The user comes into the scenario immersed, but it lacked a little awareness of what was going on in the scene surrounding the user.” This is because the user does not see any virtual environment until they choose a scene and enter it. There is no expectation for what 3D content will look like or how they will interact with 3D content. Future developers could avoid this confusion by adding an introduction into the scenario that briefs users on the mission context or by providing 3D objects within the lesson menu.

Case Study 1 Lessons Learned

To increase user immersion:

- Use multiple sensory modes (sight, sound, etc.).
- Create a natural narrative and integrate user instructions into this narrative.
- Provide stylistic 3D objects for users to anchor themselves on and set future expectations (e.g., 3D objects integrated into the lesson menu or an introductory scene before the user begins completing tasks).

To reduce cognitive load:

- Use familiar and direct interactions.
- Give users time to digest their new virtual environment before giving complex tasks.
- Provide users with direct goals for their actions (either through visual and/or verbal cues) to help direct them on what to do next.
- When giving instructions to the user, consider what is visible in their FOV while listening to those instructions. Consider highlighting features that focus on their task while making other irrelevant features less noticeable.

To maximize safety:

- If users are required to interact with the real environment (including walking around), make sure that they can easily view their real environment to be able to complete their tasks efficiently and safely.

Case Study 2: Air Force Communications and Maintenance Operations Training (Mobile – Tablet)

Information About Application

The Air Force Communications and Maintenance Operations training application (see Figure 2) was developed for the 338th Training Readiness Squadron at Keesler Air Force Base (AFB) on AR-enabled tablets, both Android and iOS. The purpose of this application is to train communication team members in the setup of the BlueSky antenna tripod. The lesson covered the steps for constructing the tripod base which the trainees could review and build through XR activities to build familiarity with the equipment that was not accessible in the classroom. This application provided the user with detailed instructions along with XR practice. By deploying this on AR-enabled tablets, this application encourages users to embody the procedural skills that they learn in a virtual environment and allows for this to be deployed and accessed easily across a variety of users.

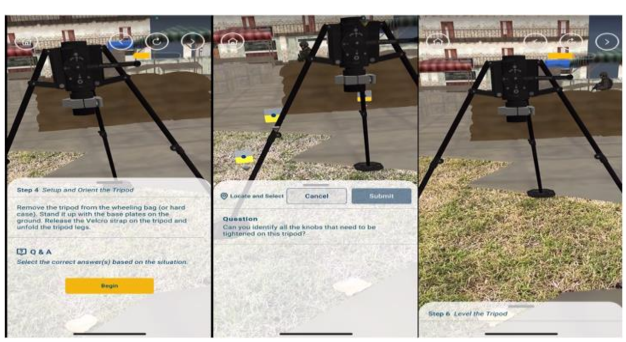

Figure 2. Air Force Communication and Maintenance Operations Application

Takeaways from the Heuristic Evaluations

Strength 1: Easily Learnable Interactions

The first strength that evaluators of the Air Force Communication and Maintenance Operations application identified was that the interactions in XR were easy to learn. Learnable interactions are crucial in XR because they enhance user engagement and usability (Hafsa et al., 2021). Unlike traditional interfaces, XR integrates digital information into the physical world, creating a more immersive and intuitive experience. When interactions are learnable, users can quickly understand and adapt to the XR environment, which reduces the cognitive load and makes the technology more accessible (Sweller, 1994). While the application did not provide an explicit tutorial, the learnability of the interactions during immersion was noted by all evaluators. This ease of learning is particularly important as XR applications become more complex and widespread, by designing interactions that users can easily learn and remember, developers can ensure a smoother, more enjoyable UX that encourages repeated use and deeper engagement with XR content.

Strength 2: Enabling Embodiment for Psychomotor Learning

The second strength identified in the heuristic evaluations for this case study was the use of embodiment in learning a psychomotor task. Evaluator E noted that "as you approach the interaction space the bars fill up to let you know when you're close enough to interact, so I had to get close to the items to learn how to do the actions." This indicates that the system is using embodied learning to encourage users to move around the space and adjust their proximity to conduct the tasks. Embodiment is essential for psychomotor learning, particularly in XR, because it leverages the natural integration of physical movement and cognitive processes (Lacey et al., 2022). In the context of XR, embodiment allows users to interact with digital elements in a way that mirrors real-world actions, enhancing the learning process. When learners physically engage with virtual objects through gestures and movements, they create stronger connections between their actions and the outcomes. This hands-on approach facilitates muscle memory development and improves the retention of complex motor skills. When developing XR applications, for both mobile and HWD, developers should encourage users to embody interactions. By simulating real-world tasks within an XR environment, learners can practice and refine their psychomotor skills in a safe and controlled setting, leading to more effective and efficient skill acquisition. This result indicates that embodiment can be successfully instantiated in XR for both mobile and HWD form factors.

Strength 3: Avoids Clutter

Another strength identified was the avoidance of clutter which is critical for usability in XR because it ensures that users can focus on the essential information and tasks without being overwhelmed. In XR, where digital elements are overlaid onto the real world, an excessive amount of visual or interactive content can lead to cognitive overload (Ferrer et al., 2013). This overload not only hinders the user's ability to process and understand the information but also detracts from the overall UX. Clear and concise presentation of information helps maintain user engagement and allows for smoother navigation within the XR environment. This application includes an instructional "drawer" which can be minimized during XR interaction and reopened when instructional text is needed, thus reducing clutter and providing an efficient and seamless interaction for the user. By minimizing clutter like in this example, designers can ensure that users can easily locate and interact with key elements, improving both the efficiency and effectiveness of their interactions.

Area of Improvement 1: Provide Detailed Help and Support

Help and support are fundamental concepts in usability because they ensure that users can effectively and efficiently navigate and utilize an application, particularly when they encounter difficulties or need guidance (Joyce, 2020). The integration of comprehensive help and support features is crucial to accommodate users' varying levels of expertise and familiarity with the technology. This application could benefit from providing a brief tutorial on the first run of the application, because as Evaluator D stated, "I was confused during one activity, but there was no 'help' section to guide me." Effective help and support mechanisms, such as contextual tips, tutorials, and real-time assistance, empower users to overcome challenges and maximize the application's potential. Additionally, the application's instructions were not detailed enough. For example, Evaluator E stated, "When you need to be 1m away and at the ground level, just saying to 'move closer' isn't enough. It would be helpful to know that you needed to bend down also." Future iterations of this application should work to improve the clarity and precision of the interaction instructions. By designing help and support features that are seamlessly integrated and responsive to users' needs, XR developers can create more user-friendly applications that facilitate learning, improve performance, and foster a positive UX.

Area of Improvement 2: Balance Consistency and Flexibility
Consistency is a crucial concept in usability because it ensures that users can predict how the application will behave, making interactions more intuitive and efficient (Krause, 2021). In XR applications, maintaining consistency across various elements such as interface design, interaction patterns, and feedback mechanisms helps users build a mental model of how the system operates. This application could be improved by ensuring that interactions and instructions are consistent throughout the experience. Allowing users to pause the application and resume at the same point in the lesson would address Evaluator F's comments that "Exiting the app will make you restart." Additionally, some evaluators noted that the virtual elements were not sized appropriately, that is, they needed to back up in the physical environment to see all of the XR content. To improve on the flexibility of the app, designers could incorporate features such as adjustable settings, context-sensitive help, and adaptive interfaces that respond to user behavior. By balancing consistency with flexibility, XR applications can provide a more inclusive, user-friendly experience that meets the diverse needs of the end-users while maintaining a coherent and predictable interaction framework.

Area of Improvement 3: Clarity of Interaction
Clarity of interactions is also critical in usability because it ensures that users understand how to interact with an application effortlessly and efficiently, and while this application's interactions were easily learnable, evaluators indicated that they were not always immediately clear. In XR, where digital elements are superimposed on the physical world, clear interactions help prevent confusion and errors, enabling users to accomplish their tasks with confidence (Cao et al., 2023). When asked if object manipulation worked well, Evaluator F responded "The 'leveling' activity where you tilt the tablet was not clear to me at all." This indicates that the interaction patterns which users are already accustomed to from smartphones and tablets were not translating directly to this activity, even though the form factor in this case was an AR-enabled tablet. Developers can avoid encountering this issue by providing more familiar and direct interactions with virtual objects, and by giving clear indicators which virtual objects can be manipulated and which cannot.

Case Study 2 Lessons Learned
To increase user help and support:

- Provide a brief tutorial on the first run of the application, as well as allowing users to access the tutorial again if they need it.
- Ensure instructions are clear and precise (e.g., if an interaction requires the user to be in close proximity to an object, ensure the instructions indicate not only distance horizontally, but vertically as well).

To maximize consistency and flexibility:

- Allow users to pause lessons and return to the same point.
- Ensure the application reminds users to be aware of surroundings when navigating in their real space.
- Use language the user will understand when providing instructions (e.g., some users may not understand terminology like "align the model").

To reduce cognitive load and clarify interactions:

- Use familiar and direct interactions to manipulate virtual objects.
- Give users clear indications of which virtual objects may be interacted with to reduce frustration.
- Provide users multimodal cues when appropriate to help direct their actions.

Case Study 3: XRMentor® (HWD – Magic Leap 2)

Information About Application
XRMentor® (see Figure 3) was developed to provide instructor-led and self-guided training for the trucking industry and was evaluated using the Magic Leap 2 HWD. In this example, XRMentor® was used to train maintenance personnel how to complete preventative maintenance tasks such as inspecting brake pads, rotors, and headlamps. This application steps the user through how to perform these maintenance tasks. By deploying this on an XR HWD, this application encourages users to embody the procedural skills that they learn in a virtual environment while providing hands-free interaction.

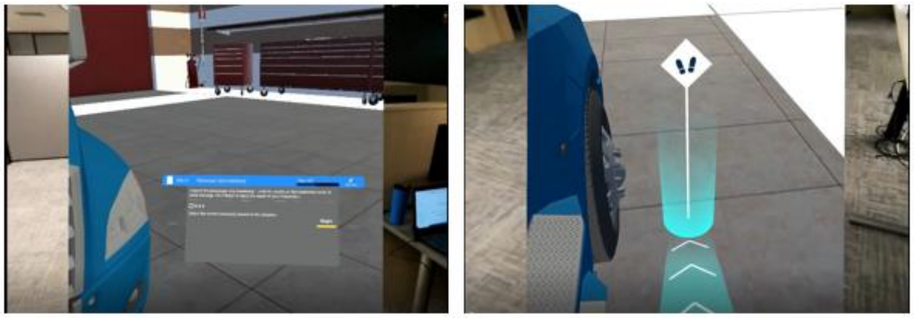

Figure 3. XRMentor® Application

Takeaways from the Heuristic Evaluations

Strength 1: Easily Learnable Interactions

Similar to findings from the Air Force Communication and Maintenance Operations application, the heuristic evaluations indicated XRMentor® was easy to use and interactions in XR were easy to learn. This is critical for XR training applications to accomplish as it enhances the usability of the application, so users spend less time learning how to use the application and more time learning the training content. XRMentor®'s interactions were easily learnable because they used common practices that users recognized with other interactions in their day to day lives. For example, they walked towards virtual objects that they needed to get a closer look at, pointed the controller to select objects (similar to a mouse cursor), and pushed the trigger button to select objects (similar to a video game controller). By providing users with interactions that are familiar, XR developers can create applications that are approachable and easy to learn.

Strength 2: Clear Aesthetics that Have a Focus on XR Content

This application included a variety of visual information in the user's FOV, including: 2D menus with text instructions, 3D environmental holograms, and 3D UI elements (see Figure 3). 2D menus are often used in XR for showing text-based instructions as they are more readable than 3D text. However, 3D holograms, which are displayed clearly and realistically using the Magic Leap 2 form factor, provide more contextual information and help the user embody the content that they are learning. By balancing these elements, XRMentor® gave a clear look and feel that did not clutter the FOV and prompted users to direct their attention to certain areas. As Evaluator G pointed out, "the step panel moves to the part of the truck where the next inspection is," which prompted them to direct themselves towards that area easily. In situations where the step panel could be occluded by the 3D truck element, a 3D walking path was placed for the user to follow to the next location. These aesthetic choices not only decluttered the FOV, but also helped orient the user to where they needed to be to continue forward in the application.

Area of Improvement 1: Provide Actionable Feedback when Users Need Assistance

During the heuristic evaluation, evaluators identified that interactions were easy to learn, and it was easy to navigate throughout the application. However, in the event a user did become confused, there was no help section to assist them. If a user were at a point where they could not continue forward, this could be frustrating. Additionally, as Evaluator H pointed out, when answering multiple choice questions in the application that quizzed their knowledge, "The system only told me if I got the question right or wrong, it did not tell me why I got it wrong or how I could been better set up to get it right". To deeply understand the learning content and improve, the user would have to find a secondary source for this information. When creating XR applications, it is important to include a way for the user to find help when they need it; and when these applications involve training content, explanations should be provided to further a user's understanding of the training material.

Area of Improvement 2: Easing the User into the Environment

Multiple evaluators noted that when they began the application, they loaded into the cab of the truck. This was jarring, disorientating, and made finding where to go next a difficult experience. The first moments in an XR application set the tone for the rest of the experience, so it is important to set the user up for success by easing them into the virtual environment for the first time. This could have been done more effectively by placing the user's start location on the outside of the truck and instructing them to look at the inside of the vehicle if it was necessary for the task to be completed.

Case Study 3 Lessons Learned

To make interactions easy to learn:

- Use familiar and direct interactions when available so users can recognize familiar actions to use the XR application.

To reduce cognitive load:

- Use a variety of techniques to show the UI within the screen space and ensure that they are located near the objects that they reference.
- Ease the user into the environment at the start of an XR application to avoid jarring and confusing the user.

To increase user help and support:

- Provide an avenue for users to ask for help if they are stuck.
- If the XR application quizzes users on their knowledge and provides feedback about whether the answer was correct or wrong, add details about why it was correct or wrong so the user can further understand the information.

DISCUSSION

This research utilized a heuristic evaluation methodology to assess the usability of three different XR applications. As a result, the researchers have identified what went well during the evaluation process and what could have been improved when conducting heuristic evaluations for XR products. These have been synthesized into actionable recommendations for future evaluators. Heuristic evaluations offer several benefits during the design of XR applications, including early identification of usability issues, cost-effectiveness, and rapid feedback. This comprehensive assessment method gathers useful user insights, ensures that usability standards are met, supports iterative design, and fosters collaboration among the development team. Early issue resolution can reduce development time, leading to higher user retention and satisfaction in the competitive XR market. Even though there are benefits to conducting heuristic evaluations, there is a lack of research in the literature citing its use within XR. In a recent systematic literature review conducted on AR/MR design heuristics, only 89 papers identified design heuristics between the years of 2000 and 2020 (Krauß et al., 2021). The lack of examples and use of heuristics and design guidelines in the literature has been identified as a key barrier in authoring AR/VR applications (Ashtari et al., 2020). This current research expands the current literature by providing three case study examples of XR heuristic evaluations. The following recommendations were identified for the successful conduct of heuristic evaluations for XR application.

Recommendations for Conducting Heuristic Evaluations

- Use 3-5 evaluators for each application (Nielsen, 1994).
  - During this study, this was beneficial because a variety of evaluator perspectives created a more comprehensive evaluation.
- Utilize a validated heuristic checklist, such as The Derby Dozen: 12 Usability Heuristics for AR and MR (Derby, 2023).
  - During this study, the chosen checklist provided an organized, effective, and straightforward approach for conducting the evaluations. It provided evaluators with a standardized form to evaluate the application while also providing the opportunity for detailed feedback. This checklist afforded creation of a comprehensive list of standard practices for XR usability design that may not otherwise have been considered, and provided data in a format that was easy to aggregate for analysis.
- Carefully consider evaluators qualitative comments in addition to checklist ratings.
  - If a heuristic checklist item is rated as a “no, this checklist item is not satisfied” this may be because it is out of the scope of the application rather than a major issue that needs to be fixed. Evaluators’ qualitative feedback provides a more robust understanding of the criticality of the rating.
- Carefully consider XR form factor during the evaluation phase.
  - XR applications can be deployed on a variety of different form factors (e.g., tablets and HWDs). In these three case studies, common strengths and weaknesses were found across applications, each of which used a different XR device. This shows that while XR usability challenges may be similar across some applications, some of the differences may be attributed to hardware. There is an

opportunity for additional research to explore the same application across different devices to isolate the form factor variable.

The heuristic evaluations reported in this paper revealed several strengths and areas for improvement in XR interaction design that future developers can utilize when designing XR applications. Firstly, the XR applications excelled at increasing user immersion by using sensory elements like sound and natural narrative and directing users to interact physically with the environment. They also effectively reduced cognitive load by employing familiar interactions and clear, linear instructions. Additionally, the balance between virtual and real-world elements helped users remain aware of their surroundings while engaging in immersive experiences. However, improvements are needed in directing users' focus, setting expectations early, providing detailed help and support, balancing consistency with flexibility, and ensuring clear interaction instructions. These insights highlight the importance of integrating sensory elements, easing users into new environments, and providing clear, familiar interactions to enhance the usability and immersion of XR applications. Moreover, this research generated several recommendations that can be applied to future XR application development. These insights confirm previous findings in the XR literature but furthers these findings to provide actionable steps XR developers can take to minimize the usability challenges experienced in their applications and increase user engagement and adoption.

Recommendations for XR Application Design

Increase User Immersion:

- Utilize multiple sensory modes by integrating sight, sound, and other senses to create a rich, engaging experience.
- Create a natural narrative by seamlessly incorporating user instructions into a cohesive story.
- Anchor scene with 3D objects by using stylistic 3D objects in menus or introductory scenes to set user expectations and guide them.

Increase User Help and Support:

- Offer brief tutorials on the first run and allow users to revisit them as needed.
- Provide precise instructions, including horizontal and vertical distances if required for interactions.
- Create avenues for users to seek help if they are stuck.
- Provide feedback with explanations when quizzing users. Offer detailed feedback explaining why answers are correct or incorrect.

Maximize Consistency and Flexibility:

- Allow users to pause and resume lessons from the same point.
- Remind users to be aware of surroundings when navigating in real space.
- Use clear language the user will understand when providing instructions and avoid technical jargon (e.g., "align the model").

Reduce Cognitive Load and Clarify Interactions:

- Employ familiar and direct interactions that users recognize and understand to manipulate virtual objects.
- Allow users to acclimate to their environment through exploration before presenting complex tasks.
- Provide clear multimodal cues to guide users' actions and focus their attention on relevant features.
- Ensure instructions are contextual and consider what is visible in the user's FOV to highlight relevant features.

Conducting heuristic evaluations provide developers with concrete guidelines for design and development of innovative XR applications. This methodology allows users to systematically explore the interface and identify strengths and weakness during the design phases while they can easily be addressed, reducing cost, and increasing efficiency of development. This type of hands-on interaction with the application helps to highlight critical usability problems that might not be apparent through other evaluation methods.

CONCLUSION

XR technology has been shown to be an effective training tool that encourages embodied learning. This technology has been advancing rapidly, and to keep improving its effectiveness, it must have strong usability. To enable this, as a community we need to identify best practices within this domain. This can be achieved through the development and use of heuristic evaluations which will identify key aspects of XR applications and hardware that lead to the improvement of overall usability. This paper contributes to this work by discussing the experience of conducting XR heuristic evaluations on three applications across multiple form factors, lessons learned to practitioners who want to conduct their own evaluations, and techniques that can be implemented during the development of XR applications to increase their usability. By providing examples of heuristic evaluations and guidelines for XR, this can break a key barrier for the development of XR applications by providing this information to XR authors and developers, and further validate current design guidelines.

REFERENCES

Ashtari, N., Bunt, A., McGrenere, J., Nebeling, M., & Chilana, P. K. (2020, April). Creating Augmented and Virtual Reality Applications: Current Practices, Challenges, and Opportunities. In *Proceedings of the 2020 CHI conference on human factors in computing systems,* 1-13.

Auganix. (2023, May 10). *U.S. Space Force training service members with XR*. Auganix.Org. https://www.auganix.org/xr-news-us-space-force-training-service-members-with-xr-suite-from-charles-river-analytics/

Augmented Reality for Enterprise Alliance. (2023). AR Heuristics Tool. *AREA*. https://thearea.org/free-ar-heuristics-tool/

Cao, J., Lam, K.-Y., Lee, L.-H., Liu, X., Hui, P., & Su, X. (2023). Mobile Augmented Reality: User Interfaces, Frameworks, and Intelligence. *ACM Computing Surveys*, *55*(9), 189:1-189:36. https://doi.org/10.1145/3557999

Congressional Research Service. (2022). *Military Applications of Extended Reality*. CRS Reports. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://apps.dtic.mil/sti/pdfs/AD1157210.pdf

Derby, J. (2023). *Designing Tomorrow's Reality: The Development and Validation of an Augmented and Mixed Reality Heuristic Checklist* [Dissertation, Embry-Riddle Aeronautical University]. https://commons.erau.edu/cgi/viewcontent.cgi?article=1794&context=edt

Ferrer, V., Yang, Y., Perdomo, A., & Quarles, J. (2013). Consider your clutter: Perception of virtual object motion in AR. *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 1–6. https://doi.org/10.1109/ISMAR.2013.6671835

Hafsa, S., Majid, M. A., & Tawafak, R. M. (2021). Learnability factors of AR usage performance: Validating through survey. *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM),* 371–376. https://doi.org/10.1109/ICSECS52883.2021.00074

Hopkins, B. (2021, September 29). *Forrester's 2021 Technology Trends*. Forrester. https://www.forrester.com/report/forresters-2021-technology-trends/RES176252

Joyce, A. (2020, December). *Help and Documentation (Usability Heuristic #10)*. Nielsen Norman Group. https://www.nngroup.com/articles/help-and-documentation/

Krauß, V., Jasche, F., Saßmannshausen, S. M., Ludwig, T., & Boden, A. (2021, December). Research and practice recommendations for mixed reality design–different perspectives from the community. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, 1-13.

Krause, W. L. in R.-B. U. (2021, January). *Maintain Consistency and Adhere to Standards (Usability Heuristic #4)*. Nielsen Norman Group. https://www.nngroup.com/articles/consistency-and-standards/

Lacey, G., Gozdzielewska, L., McAloney-Kocaman, K., Ruttle, J., Cronin, S., & Price, L. (2022). Psychomotor learning theory informing the design and evaluation of an interactive augmented reality hand hygiene training app for healthcare workers. *Education and Information Technologies*, *27*(3), 3813–3832. https://doi.org/10.1007/s10639-021-10752-4

Nielsen, J. (1994, November 1). *The Theory Behind Heuristic Evaluations, by Jakob Nielsen*. Nielsen Norman Group. https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/theory-heuristic-evaluations/

Rasmussen, A. (2023, October 4). *AFSOC embraces Extended Reality (XR) to enhance readiness*. Air Force Special Operations Command. https://www.afsoc.af.mil/News/Article-Display/Article/3548421/afsoc-embraces-extended-reality-xr-to-enhance-readiness/https%3A%2F%2Fwww.afsoc.af.mil%2FNews%2FArticle-Display%2FArticle%2F3548421%2Fafsoc-embraces-extended-reality-xr-to-enhance-readiness%2F

South, T. (2023, September 7). *Army approves next phase for augmented reality device*. Army Times. https://www.armytimes.com/news/your-army/2023/09/07/army-approves-next-phase-for-augmented-reality-device/

Stanney, K. M., Skinner, A., & Hughes, C. (2023). Exercisable Learning-Theory and Evidence-Based Andragogy for Training Effectiveness using XR (ELEVATE-XR): Elevating the ROI of Immersive Technologies. *International Journal of Human–Computer Interaction*, *39*(11), 2177–2198. https://doi.org/10.1080/10447318.2023.2188529

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5