Fall 2023

Designing Tomorrow's Reality: The Development and Validation of an Augmented and Mixed Reality Heuristic Checklist

Jessyca Derby
Embry-Riddle Aeronautical University, derbyj1@my.erau.edu

Designing Tomorrow's Reality: The Development and Validation of an Augmented and Mixed Reality Heuristic Checklist

Jessyca L. Derby

B.A., Psychology, Keene State College, 2018

M.S., Human Factors, Embry-Riddle Aeronautical University, 2020

A dissertation submitted to the Department of Human Factors and Behavioral Neurobiology in the College of Arts and Sciences in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Human Factors.

Department of Human Factors and Behavioral Neurobiology

Embry-Riddle Aeronautical University

Daytona Beach, Florida

October 2023

Acknowledgements

First, I want to thank my advisor and dissertation chair, Dr. Barbara Chaparro. Thank you for your guidance all these years. You were always there when I needed a brainstorming session just to get the ideas out, and always helped me persevere when times were tough. You provided me with so much unmatched guidance that allowed me to grow into the researcher I am today. Thank you for all your support.

Similarly, thank you to my committee members (Dr. Alex Chaparro, Dr. Christina Frederick, Dr. Shawn Doherty, and Dr. Jonathan Kies), and the Augmented Reality for Enterprise Alliance (AREA) for their direction and support throughout this process. Each of your unique expertise has made this heuristic checklist what it is today. And thank you to my undergraduate advisor, Dr. Martin Brown who pushed me to pursue the type of research I felt passionate about.

A special thank you to all of those who have helped contribute to this process. This heuristic checklist could not have been made without those who volunteered their time to be expert reviewers, heuristic evaluators, and user testing research assistants. You all are just as excited to see this checklist become a reality as much as I am, and I cannot wait to distribute it out to the research world with you all.

I could not have gotten through this dissertation without the support of my friends and family. To my mother who has always been there with her support. She knew from the start that I could do this. My stepfather, who is there for me to talk to about work, life, and everything in-between. My brother, who has always been there for me to provide advice, guidance, or to just listen to me complain. Paul, in your dissertation acknowledgements you told me that, “education opens a world of opportunity if you’re willing to dedicate yourself to it for a relatively small number of years”. And I ask, “a relatively small number of years”? I have been a student my entire life, and that may never change. But, wow, I do ever value and love it. To my father, who instilled in me the importance of curiosity and the pursuit of knowledge. And to my grandparents who are always excited to hear about my progress, as they know how difficult this journey truly is.

Thank you to my support system at Embry-Riddle: Jacqueline Almada, Emily Anania, Cassandra Domingo, Carmen Van Ommen, Logan Gisick, John Kleber, William Shelstad, Emily Rickel, and many others. Thank you for the late-night study groups, the board game nights, and our venting sessions. You may be my colleagues, but you are also lifelong friends. And, finally, to my partner in crime, Kyler. You are my support, my rock, and without you always bringing me back down to earth, this process would have probably consumed me.

I do not know what I would have done without all of you cheering me on from the sideline. Thank you all.

“I’VE GOT BLISTERS ON MY FINGERS!”

- Ringo Starr

Abstract

Augmented (AR) and Mixed Reality (MR) are new and currently developing technologies. They have been used and shown promise and popularity in the domains of education, training, enterprise, retail, consumer products, and more. However, there is a lack of consistency and standards in AR and MR devices and applications. Interactions and standards in one application may drastically differ from another. This may make it difficult for users, especially those new to these technologies, to learn and feel comfortable using the devices or applications. It may also hinder the usability of the applications as designers may not follow proven techniques to display this information effectively. One way to create these standards is through the development and acceptance of usability or user experience (UX) heuristics. There is a lack of validated and widely accepted heuristics in AR and MR. Those that do exist tend to be too specialized to be valid across types of applications or devices. This dissertation's goal is to fill this gap through the creation of a validated usability/user experience (UX) heuristic checklist to evaluate AR or MR devices and/or applications by following a validated methodology for developing usability/user experience heuristics (Quiñones et al., 2018).

Previous work had been completed to develop an AR and MR heuristic checklist (Derby & Chaparro, 2022). This work resulted in 11 heuristics and 94 checklist items; however, validation of this checklist was limited. This dissertation broadened the heuristic checklist to ensure applicability to more application types, device types, and use cases. Five different applications and devices were used to validate the checklist through heuristic evaluations and user tests. Experts in the domain also provided their feedback on the heuristic checklist using applications of their choice. A total of 100 revisions were made to the Derby & Chaparro (2022) checklist as a result of this study. The final heuristic checklist consists of 12 heuristics and 109 checklist items that practitioners can use to evaluate AR or MR applications and devices and quantify the results to better inform design.

Keywords: augmented reality, mixed reality, user experience, heuristics, best practices

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction & Background

Purpose Statement

The purpose of this dissertation is to develop and validate a usability/user experience (UX) heuristic checklist that can be used to evaluate any Augmented Reality (AR) or Mixed Reality (MR) device and/or application. Previous work had been completed to develop this heuristic checklist (Derby & Chaparro, 2022). This work resulted in 11 heuristics and 94 checklist items. Appendix A lists the heuristics and checklist items. Results from the validation of this checklist showed that further improvements needed to be made to cover important usability considerations for AR and MR. As described in the literature review of this document, there is a variety in the types of AR and MR technology and applications. There are head-mounted displays, mobile hardware, projection-based technology and all are being used in a variety of domains (education, training, medical, maintenance, entertainment, retail, etc.). The heuristic checklist that was developed previously was only validated with two AR and MR applications (a medical education application, and a puzzle game), and three types of devices (mobile phone, the Magic Leap 1, and HoloLens 1) (Derby & Chaparro, 2022). None of which included technology like head-mounted video see-through (VST) displays that are becoming more popularized in MR devices like the Apple Vision Pro and Meta Quest 3 (Apple, 2023a; Meta, 2023). The validation identified additional usability issues that may impact the user's experience when using AR or MR technology. For example, the heuristic checklist included items related to collaboration in AR and MR, but validation was not completed with applications that included multiple users because there were very few applications publicly available with these features. The validation only included mobile devices and two head-mounted displays (the Magic Leap 1 and HoloLens 1). Usability aspects related to comfort and inclusivity of hardware design could differ for devices such as lightweight smart glasses or video see-through (VST) displays. Additionally, aspects related to user and bystander privacy were not adequately evaluated during this validation process since the applications chosen did not require data from the user or for the user to walk around in public while using the application. This dissertation allowed for informed broadening of the heuristic checklist to insure generalizability to more application types, device types, and use cases. The final product of this dissertation resulted in a toolkit that practitioners can use to evaluate AR or MR applications and devices and quantify the results to better inform design.

Background

The concept of augmented reality (AR) and mixed reality (MR) is not new. Definitions of this technology have been noted as far back as 1968 (Sutherland, 1968). However, this technology has been improving at a rapid rate since then. AR and MR technologies in the early 1970s, or even as recent as the 2000s, are very different than what we see today. Yet, there is still a lack of agreed upon best practices for the design of these technologies. Researchers and companies in this space have begun to establish their own best practices, but there is a lack of validated usability heuristic checklists that can be used across device types and applications. This gap can result in difficulties for those who are developing AR and MR applications, especially those who plan to deploy the application on multiple types of devices (e.g., mobile devices and head-mounted displays (HMDs)). During the development process, professionals can test their prototypes by conducting user tests and receiving feedback from their potential users, but this can be time consuming and require many resources. Heuristic checklists allow professionals to evaluate an application or device without requiring end users and can even be done early in the prototyping process. Heuristic evaluations can provide a starting point for designers and developers to iterate on their designs to make the technology more usable and elevate the user's experience. The development of an AR and MR usability heuristic checklist would be beneficial to the development and design of these technologies.

Previous work had been completed to develop this AR and MR heuristic checklist (Derby & Chaparro, 2022). The authors created a heuristic checklist with 11 heuristics and 94 items by following an eight-step methodology for developing and validating usability heuristic checklists (Derby & Chaparro, 2022; Quiñones et al., 2018). Appendix A lists the heuristics and checklist items. Results from the validation of this checklist showed that further improvements needed to be made to cover important usability considerations for AR and MR. The work conducted in this current study was necessary to further validate and improve this heuristic checklist.

Chapter 2

Literature Review

Introduction

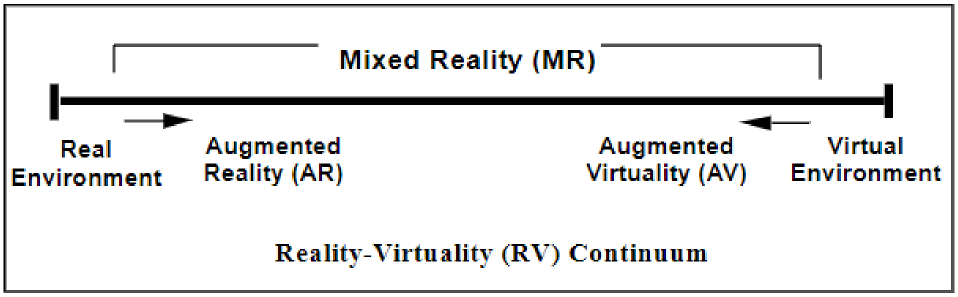
This chapter provides a review of relevant literature. This includes definitions, rationale for the areas considered in the heuristic checklist based on human perception and usability/UX standards, and additional published literature regarding AR and MR.

Defining Augmented and Mixed Reality

The first AR head-mounted display (HMD) is attributed to Ivan Sutherland in 1968, however the term was not commonly used until Thomas Caudell, a Boeing researcher, coined it in the 1990s (Sutherland, 1968; Caudell & Mizell, 1992). Throughout the years, the definition of AR has changed as technology has changed and improved. Caudell explains AR as a heads-up display head set that, "augment the visual field of the user with information necessary in the performance of the current task) (Caudell & Mizel, 1992). Later, Milgram et al. defined AR using a Reality-Virtuality (RV) Continuum shown in Figure 1 (1994). The left side of the continuum represents the real environment, where the environment, objects, and information that users see and interact with are part of the real world. The right side of the continuum represents a fully virtual environment where the environment, objects, and information are all created through computer graphics, as in Virtual Reality (VR). AR is near the left side of the continuum, showing that some virtual elements are included but mostly the real environment around a user stays the same, and the authors define it generally as, "augmenting natural feedback to the operator with simulated cues" (Milgram et al., 1994). This allows for both head-mounted displays and monitor-based displays to be included in this definition.

Figure 1

Visualization of Milgram's Reality-Virtuality (RV) Continuum (1994)



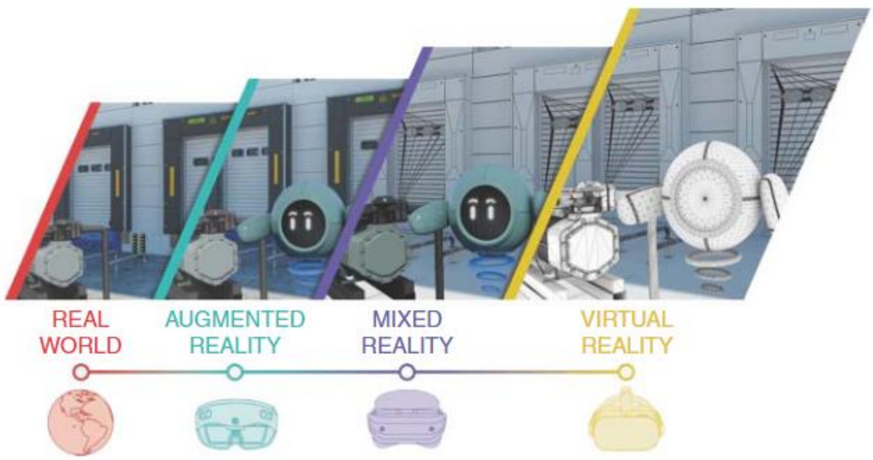
Note. Image from Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1994). Augmented reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies, 2351*, 282-292. https://doi.org/10.1117/12.197321

Milgram et al. (1994) also propose a definition of Augmented Virtuality (AV) which is similar to AR, however, the environment is either partially or entirely virtualized. An example of AV: a user is in a fully computerized virtual environment but a real object, such as their hands, are introduced to grab something in the virtual world (1994). These authors also define MR, which spans the entire continuum, and is defined as an "environment as one in which real world and virtual world objects are presented together within a single display" (Milgram et al., 1994). With this definition, AR or AV HMD-based displays are also considered MR displays. Other definitions later emerged to avoid specifying a type of display, as AR displays were constantly innovating.

Now not only do we have HMD and monitor-based AR, but also AR on mobile phones, projector-based AR, and audio-only AR. Azuma (1997) used the following definition in order to take the display out of the definition, "AR [are] systems that have the following three characteristics: 1) Combines real and virtual 2) Interactive in real time 3) Registered in 3-D" (Azuma, 1997). Other definitions completely separating AR and MR have become more commonplace as technology has developed. These definitions differentiate AR and MR through the actions of simply “overlaying” information onto the real world and “combining” the virtual and real elements through computer vision so that the virtual content is aware of and interacts with the real-world environment (Lenovo, n.d.; Stanney et al., 2021). For example, virtual objects such as holograms act differently on an AR device like Google Glass and a MR device like the Microsoft HoloLens. Though they are both see-through displays that a user puts over their eyes, Google Glass overlays virtual information on top of the physical environment whereas the information that the Microsoft HoloLens overlays can interact with the environment around it (following the rules of physics by scanning the surfaces of the environment and letting objects fall, roll, and sit based on where it is placed on those surfaces). Stanney et al. created a visualization of this updated virtuality continuum, and it is shown in Figure 2 (2021). Extended Reality (XR) is also being used more recently as a term that describes everything on this spectrum (AR, MR, VR, etc.) (Stanney et al., 2021).

Figure 2

Visualization of Stanney's Virtuality Continuum (2021)

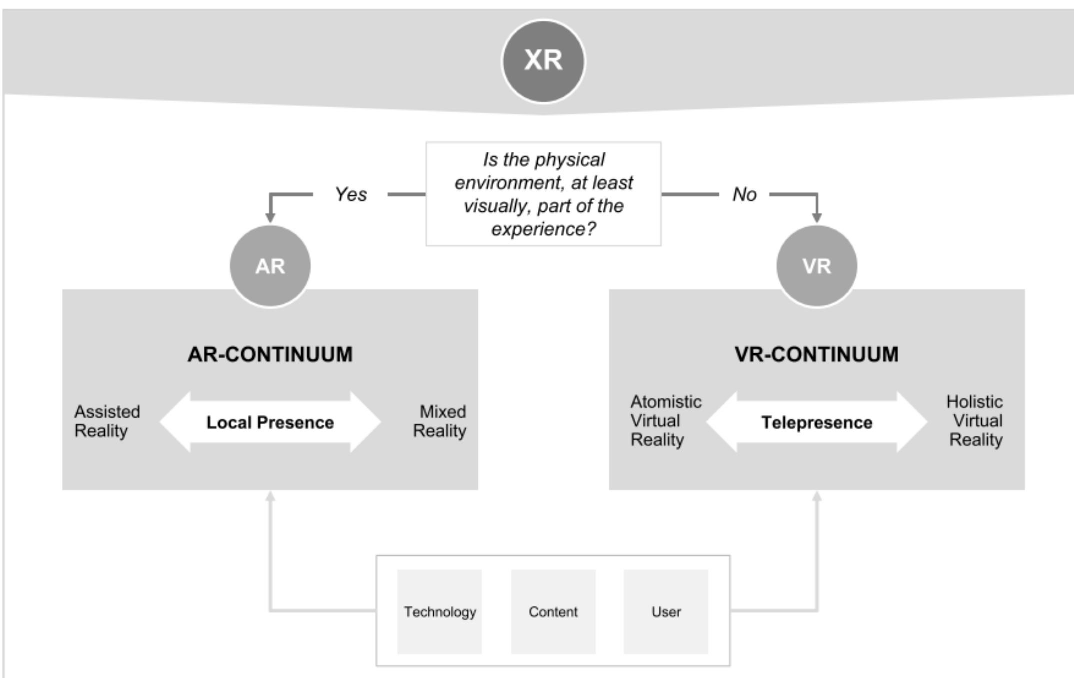


Note. Image from Stanney, K. M., Nye, H., Haddad, S., Hale, K. S., Padron, C. K., & Cohn, J. V. (2021). Extended reality (XR) environments. In G. Salvendy & W. Karwowski (Eds.), *Handbook of Human Factors and Ergonomics* (5th ed., pp. 434-456). John Wiley & Sons.

Recently, an additional framework has been proposed that describes less of a distinction between AR and MR (Rauschnabel et al., 2022). This framework is pictured in Figure 3. In this framework, XR is still used to describe both AR and VR. Though, instead of XR being called "extended reality" this abbreviation is used for the term "xReality". This is due to the idea that this technology does not "extend reality" but instead represents an "alternative reality". VR is still described as a representation of a fully virtual environment where the environment, objects, and information are all created through computer graphics. The definitions of AR and MR differ than what has been defined previously. In this framework, AR is used to describe a technology that combines real and virtual content that is displayed in real time (Rauschnabel et al., 2022). AR is separated from VR "based on whether the physical environment is, at least visually, part of the user experience (=AR) or not (=VR).” (Rauschnabel et al., 2022). MR is described as a type of AR and lies on one side of the AR continuum and describes an experience in AR that is seamless and “impossible to detect” as different from the real environment. The other side of the AR continuum is assisted reality and is often characterized by overlays in the real environment and feels less integrated and disconnected from the real environment (Rauschnabel et al., 2022).

Figure 3

Rauschnabel’s XReality (XR) Framework: Augmented and Virtual Reality (2021)



Note. Image from Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). What is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior, 133,* https://doi.org/10.1016/j.chb.2022.107289

AR and MR have been defined in many ways as these technologies have evolved. In this document, Stanney et al.'s definitions of XR, AR, and MR (2021) is used to define AR and MR, "eXtended Reality [XR] solutions are computer-generated immersive environments that provide a spectrum of experiences including... Augmented Reality (AR), which overlays virtualized content onto the real world; Mixed Reality (MR), which in addition to augmenting the real world with virtualized content, allows the virtual content to be aware of and interact with the real world; or vice versa with real objects in a virtual world" (Stanney et al., 2021, p.783).

What About Virtual Reality?

VR is encompassed in the definition of XR, but is functionally different from AR or MR. The environment and information are fully virtualized (Milgram et al., 1994; Stanney et al., 2021). In most cases, information is not overlayed or integrated with the real environment. Basic interactions also differ from VR, as this is primarily done with game-controllers while AR and MR interactions are primarily done through hand gestures or touch on a mobile device (Stanney et al., 2021). Since VR is functionally different from AR and MR, usability, design principles, and recommendations will differ. In fact, separate design heuristics have been developed for a variety of VR environments (Sutcliffe & Gault, 2004; Murtza, et al., 2017; Desurvire & Kreminski, 2018; Smith, 2021).

Types of Augmented and Mixed Reality

Since AR and MR are such new technologies, they are still changing and developing. As discussed above, even the basic definitions of these devices have changed throughout the years as technology has improved. Currently, there are a few specific types of AR and

MR devices on which these applications can run: mobile (e.g., a cellphone or tablet), head-mounted display (HMD) like the Microsoft HoloLens or Magic Leap, desktop and projection-based devices. These can be further classified as optic see-through (OST) or video see-through (VST) displays. OSTs use transparent lenses that allow direct views of the environment whereas VST displays a video feed to indirectly view the environment (Park et al., 2020). An example of an OST is the Microsoft HoloLens, pictured in Figure 4 (Microsoft, 2022a). The user views the virtual and real environment through transparent lenses. An example of a VST is mobile AR or MR using a smartphone device, where the user views the virtual and real environment through a camera and smartphone screen. An example of this is also pictured in Figure 4. The AR and MR applications on each of these devices may function very differently based on whether they are marker-based, marker-less, or location-based applications.

Figure 4

Examples of an HMD OST MR Device and a Mobile VST AR Application

Note. Left: The Microsoft HoloLens 2, an HMD OST MR device (Microsoft, 2022a). Right: Google Maps Live View, a mobile VST AR application (Google, n.d.a). Image on left from Microsoft (2022a). Microsoft HoloLens | Mixed Reality Technology for Business. https://www.microsoft.com/en-us/hololens

Hardware

Mobile AR and MR applications are deployed on a mobile device, such as a tablet or smartphone, using systems such as Apple's ARKit or Google's ARCore (Hillman, 2021; Stanney et al., 2022). Many of these applications are distributed to consumers ready-to-use in their application stores. Some examples include games (Angry Birds, 2022; Niantic, 2022), content creation apps (Snapchat, n.d.), retail (IKEA, 2022; Warby Parker, 2022), and educational apps (Anima Res, n.d.; BBC, 2018).

Wearable AR devices are deployed on hardware that sit on top of the user's head in front of their eyes, these are known as head-mounted displays (HMD) or glasses (Hillman, 2021; Stanney et al., 2022). Some examples include the Microsoft HoloLens 2, Magic Leap One, Google Glass, and Lenovo ThinkReality A3 Glasses (Google, n.d.b; Lenovo, 2022; Microsoft, 2022a; Magic Leap, 2022). All of the hardware is contained in the headset, processing the virtual information on its own and translating it to the user using internal sensors and spatial mapping systems. These devices are costly and are often seen in commercial and enterprise use cases rather than for a general consumer's home use.

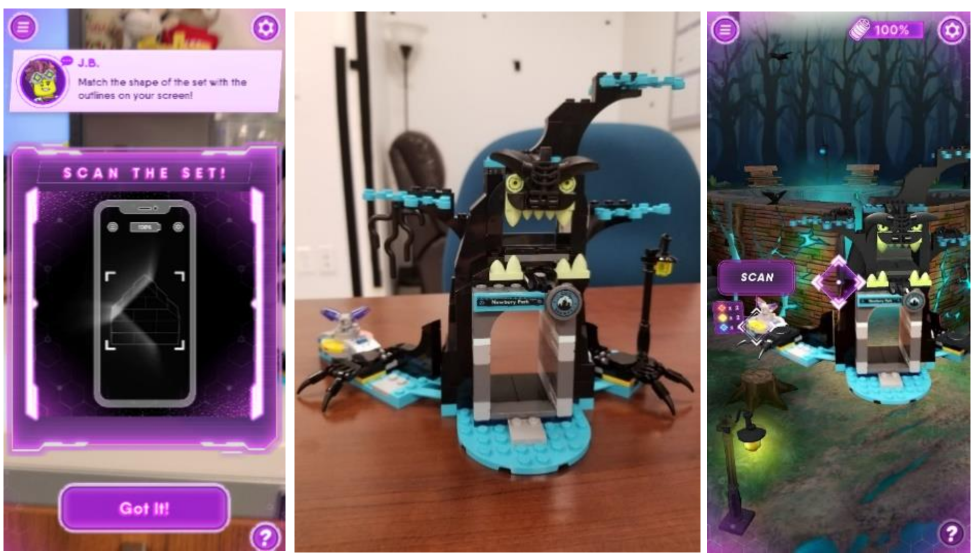
Projection-based and desktop AR and MR applications are more specialized than wearable or mobile devices. Projection-based devices involve displaying virtual information onto the real environment through one or multiple projectors, so a device is not necessary for the user to hold or wear. Some examples for this include windshield AR for cars, military sandtables, and even interactive holograms at Disney World (Hillman, 2021; Amburn et al., 2015; Mine et al., 2012). Desktop AR deploys AR experiences using a desktop screen and 3D glasses (Holo-SDK, 2020). Applications are not broadly developed for each of these technologies, and they are not as popularized as mobile and wearable AR and MR since they require a large system or specialized training to use.

Functionality

Any AR or MR application that is deployed on a mobile device, wearable, projection-based, desktop application, or any other technology may act functionally differently based on how the virtual elements are integrated into the real environment. This can be done though marker-based techniques where the virtual element is tethered to a real object. This can be as simple as a QR code or as complex as a 3-dimensional shape (Hillmann, 2021). As long as the real object is in the line of sight of the device's camera or sensor, the virtual element will appear. As the real object moves, the virtual element may move with it. Examples of marker-based applications include some Snapchat filters that require a user's face to see the objects, or LEGO Hidden Side sets that require the LEGO build to be completed and in view of a mobile device in order to use the AR features (Snapchat, n.d.; LEGO, 2022). Images of this type of application are shown in Figure 5.

Figure 5

Example of a Marker-Based AR Application

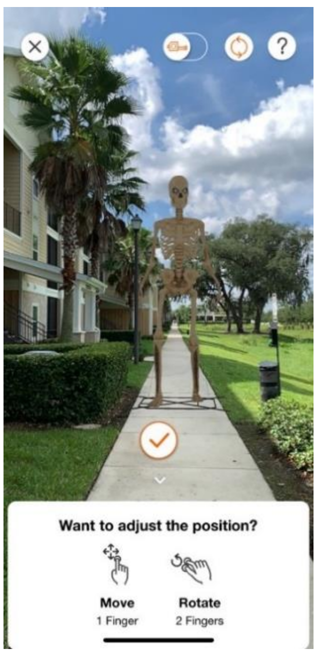


Note. The LEGO Hidden Side application requires the view of a marker (the physical LEGO set) shown in the center image. If this application sees the marker, it will overlay the virtual content on top of it, as shown in the image on the right.

Marker-less techniques use spatial localization, mapping, and computer vision to identify surfaces in the real world and place the virtual element on those surfaces (Hillmann, 2021). A specific image or object in the real world is not necessary to see these virtual elements. Examples of this include Google AR Search, Angry Birds Isle of Pigs, or retail applications such as Home Depot or IKEA Place, applications that do not require a specific object to place and view the virtual elements (Google n.d.a; Crets, 2020; Angry Birds, 2022; IKEA, 2022; Niantic, 2022). An image of this type of application is shown in Figure 6.

Figure 6

Example of a Marker-Less AR Application

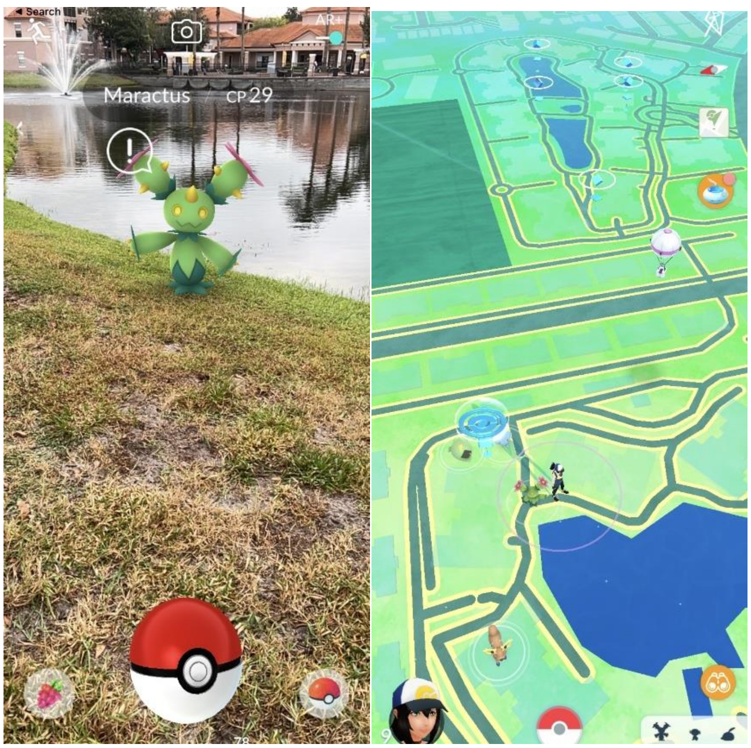


Note. The virtual content (the skeleton) in this Home Depot AR marker-less application can be moved anywhere the user places it.

Location-based techniques tether the virtual elements to a real point in space using a GPS or other positioning system in the device (Hillmann, 2021). A user is required to be physically present at a specific location in order to see and interact with the virtual content. Examples of this include Pokémon Go and Google Maps AR, and are common with wayfinding applications (Google, n.d.a; Niantic, 2022). Images of this type of application are shown in Figure 7.

Figure 7

Example of a Location-Based AR Application



Note. Pokémon Go's virtual content is tethered to a specific location in the real world. The left image shows where this virtual content is located and the image on the right shows more information if you click on virtual content but are not near it.

Perceptual Considerations of AR and MR

AR and MR devices provide virtual enhancement of an environment. This may be visual, auditory, or tactical in nature. Most devices provide visual enhancements, making visual content is the primary driver of an AR or MR experience. It is important to consider the visual experiences that users have with virtual content in order to ensure the best and most streamlined use of this virtual content. Some visual issues that may occur when using AR or MR devices include visual acuity and monocular and binocular cues for depth perception (Livingston et al., 2013; Potemin et al., 2018; Zhdanov et al., 2019; Erickson et al., 2020).

Visual Acuity

Visual acuity, the ability to see details, is crucial for the legibility of text and fine details of 3D objects in AR and MR applications. If a user cannot see details of the virtual content, they may not be able to use an application efficiently or effectively. A decrease in pixel resolution, as well as a decrease in size and contrast of virtual content have all been shown to decrease visual acuity (Livingston et al., 2013; Potemin et al., 2018; Erickson et al., 2020). Environmental lighting has shown to play a role as well. Virtual objects should be illuminated to match the real light sources of the real environment. This includes aspects of brightness, shadows, and reflections (Potemin et al., 2018). However, there tends to be a balancing act when it comes to lighting sources. A recent meta-analysis found that as lighting of the environmental increases, such as when a user is outdoors, performance tends to decrease, possibly due to the impact that increased lighting has on decreasing contrast of virtual content, causing a “washing out” effect of the virtual content (Potemin et al., 2018; Erickson et al., 2020). It is important for virtual content to match environmental lighting, but this may cause visual acuity to decrease as contrast may decrease if the lighting is too bright.

Monocular and Binocular Cues

One of the most distinguishing factors of visual AR and MR content is that some virtual content is often three-dimensional. This is often portrayed by the ability to rotate and interact with all sides of the object, and by conveying a sense of depth using monocular and binocular cues. Common monocular cues include occlusion (the act of hiding parts of one, further, object behind a closer object), relative size (as one object is closer than another, it increases in size), and shadows that display a blockage of light (Goldstein, 2014). It is recommended to follow these cues with virtual content AR and MR environments to give the user a perception of depth that matches their real-world experiences (Furmanski et al., 2002). These cues can be applied for multiple virtual objects (e.g., one closer object occluding another object that is further away) and for information about how virtual content is placed in the surrounding environment (e.g., a virtual object may be occluded by a table in the real environment if the user must move behind the table in order to interact with the virtual content). Virtual grids and distance markers are also used in order to enhance the feeling of depth (Furmanski et al., 2002).

Some devices, such as HMDs, also enable the use of binocular cues for stereoscopic vision. They enable vergence and accommodation through encouraging binocular disparity by showing slightly different images of the virtual content on each eye (Goldstein, 2014). To do this, interpupillary distance can be calculated through eye calibration. This information is used to assist with the accommodation and convergence of the eyes that occurs when an object moves closer or further away. Accurate eye calibration is extremely important for this experience to be functional. Inaccurate calibration can cause eye strain, headaches, and poor views of virtual content (Livingston et al., 2013; Park et al., 2020).

Some AR and MR applications provide more than just visual cues to the user. This may occur as auditory cues (sound effects or 360-degree audio, also known as spatial audio), or tactile cues such as vibration of the device or controllers. Multimodal interactions in XR have been shown to be beneficial to convey information as it can improve speed of task completion, capture users' attention, and increase a sense of presence (Hale & Stanney, 2015).

Auditory Cues

Even though AR and MR are heavily focused on visual information, auditory cues have been shown to enhance the experience as well. Auditory cues can increase a sense of presence, selective attention, and localization when spatial audio is used (Hale & Stanney, 2015; Stanney et al., 2021). Spatial audio can map auditory information to a location in the virtual space and can be beneficial for locating objects that require users' attention away from the current view in front of them. For example, a user may need to press a virtual button on a device that is off screen, to the left of them, to continue the process. A tone sent to the left speaker of the device could alert the user that a virtual object unseen to them is requiring their attention. According to Stanney et al. (2021) it is recommended that spatial audio should:

1) Be as similar to the real world as possible. This includes originating from the correct direction and modeling loudness with distance to enhance a feeling of presence.

2) Provide as much information from the scene as possible (the authors specifically give an example of, "the sound of a metal object interacting with another metal

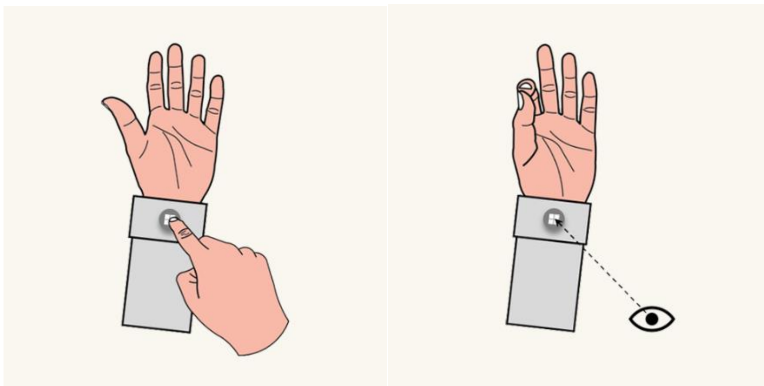
object should make a distinct sound compared to a metal object interacting with carpet). Audio can also assist with features that would traditionally be seen with haptic cues (such as clicking a button) if haptic feedback is not available.

Tactile Cues

Tactile cues portray a feeling of touch, such as vibrations and haptic feedback. In AR and MR, this type of cue often signifies that virtual content was selected, further confirming to the user that their input was processed (Hillmann, 2021). This can commonly be felt through the vibration of a mobile device or controller. Many examples of this in AR or MR are still in its infancy and not widely used. For example, some devices, such as the Microsoft HoloLens 2, are unable to provide vibrations as tactile feedback due to a lack of controllers. The HoloLens 2 can still provide some tactile feedback through hand gestures that involve touching a part of the hand or wrist. This is seen with the start gesture, which requires the user to physically touch their wrist or two fingers together, as shown in Figure 8.

Figure 8

HoloLens 2 Start Gesture



Two-handed One-handed

Note. The two-handed gesture is used to open the Start menu (left). One-handed gesture used to open the Start Menu (right). Image from Microsoft (2022b). Mixed Reality docs - Start gesture. https://docs.microsoft.com/en-us/windows/mixed-reality/design/system-gesture

Perceptual Considerations Summary

The goal for AR and MR is to develop the applications to act as the user would expect them to, making them as intuitive as possible (Hillmann, 2021; Stanney et al., 2021). This means that objects and their cues should match the real environment that they are in – lighting, shadows, and basic rules of human perception should be followed. However, one benefit of AR and MR is that it can enhance the environment. A 3D virtual model of a human heart can be beating and floating in mid-air with information surrounding it. This would be impossible to view this the same exact way in real life, but it may still make sense to the user in AR or MR due to the context and the use case. This experience has been further described as a "reality enhancing system – one that is 'transparent and more a part of the users' perceptive system than a separate entity in itself'" (Nilsson, 2010). The virtual content does not need to be a replica of the real environment but should be designed to be believable enough to co-exist with it.

User Experience (UX) and Usability

UX is the end-to-end experience a user has with a product or process. It is defined as, "user's perceptions and responses that result from the use and/or anticipated use of a system, product, or service" (International Organization for Standardization, 1998). Some factors that are included in UX are often referred to as the User Experience Honeycomb: useful, usable, desirable, findable, accessible, credible, and valuable (Morville, 2004). A visual representation of this honeycomb is shown in Figure 9 (Morville, 2004).

Figure 9

A Visual Depiction of the User Experience Honeycomb (Morville, 2004)

Note: Image from Morville, P. (2004, June 21). *User experience design.* Semantic Studios. http://semanticstudios.com/user_experience_design/

Usability is the component “usable” of this honeycomb. It is defined by how effective, efficient, and satisfied a user is when completing a task with the device or system (International Organization for Standardization, 1998). Effectiveness considers how usable the device is. This answers the question of: Can the user successfully complete the task at hand? Efficiency takes into account the number of resources the user needs in order to complete the task. Resources can include time, human effort, costs, and materials. Satisfaction refers to how the user feels after using the product or service. This includes their physical, cognitive, and emotional responses. Additional UX and usability factors include those proposed by Nielsen (learnability, efficiency, memorability, errors, and satisfaction) as well as those proposed by ISO standards of usability (effectiveness, efficiency, and satisfaction) (International Organization for Standardization, 1998; Nielsen, 2012). These 14 factors are referenced as Usability and UX attributes in Quiñones et al.’s methodology to develop usability/user experience heuristics (2018).

Usability Considerations of AR and MR

User Interaction and User Feedback

User interactions and user feedback are key for developing an effective AR or MR experience. The user must be able to interact with the content and receive feedback from the system that confirms that the input they provided was received. This feedback can also be provided without user input by conveying the system’s status to the user, as often seen with loading bars (Harley, 2018). By providing this information, the user can feel in control and trust the system (Harley, 2018). This is true for any device and AR and MR are no exceptions. Users should be aware of the system’s status (if it is processing an input, loading an application, received user input, etc.). Examples of this could include a vibration of a controller or mobile device if the user selects virtual content, or more specifically, the appearance of the Microsoft Start icon as a user begins to complete the gesture to open the Start menu as shown previously in Figure 8 (Microsoft, 2022b).

Help and Documentation

Help and documentation are common resources that can improve the usability of applications and devices. As it is described in Nielsen's usability heuristics, the best device is intuitive and does not need any explanation, but sometimes it is necessary to provide adequate help for users who get stuck (Joyce, 2020). This can be provided as either proactive, helping the user before an issue is found to get them acquainted with the interface as a tutorial, or reactive, to troubleshoot problems. AR and MR are no exceptions to this rule. Tutorials are useful to explain to new users how to complete gestures and interact with the device. It is common that users have never interacted with technology like AR and MR before. In addition to this, online help can be useful when detailed and provide actionable descriptive help.

Accurate Integration of Real and Virtual Worlds

What makes AR and MR different from other technologies is that they enhance the real environment with virtual content. This content is either overlayed or integrated with the real environment. For this to be effective, virtual content should act as users expect, as discussed in the previous section, perceptual considerations of AR and MR. This not only includes the interactions with virtual content, but the virtual content itself should act as expected. Content should not lag far behind user input or jitter/glitch as it is being moved and interacted with. In some use cases, the accuracy of 3D objects is important as well. In an application such as IKEA place where users place virtual holograms of furniture to see how it looks in their homes, it is important that the size, shape, and color of the hologram matches the real object to show an accurate representation of the piece of furniture (IKEA, 2022). Another example would include medical applications that display virtual holograms of organs over a patient or an application that overlays information about where the wiring of a building is located within the structure. In these cases, accurate placement of holograms is of upmost importance for the application to be usable.

Consistency and Standards

There is a lack of consistency between AR and MR applications (Derby & Chaparro, 2022). One application may create a function for a gesture that is different from another. This gives creative freedom to the developers but could confuse users as expectations cannot be set if there is no consistency across applications. This is also seen as one of the large contributors to applications that make sense to users, and is referenced as one of the 10 Nielsen Usability Heuristics, and increase learnability while reducing confusion (Krause, 2021).

Physical Comfort and Safety

It is important for a user to feel comfortable and safe when completing a task with an AR or MR device. As discussed in the previous section, perceptual considerations of AR and MR, these devices can cause eye strain, migraines, and discomfort to users. The weight of a device can play a role as well. If a user is required to wear a heavy device on their head, they could experience neck strain. Physical fatigue may also be caused by holding a large mobile device, such as an iPad, for an extended period of time or performing gestures with large effortful movements.

Situational awareness has also been an area of study within AR and VR. Situational awareness is, "a person's understanding of what is happening in the current situation" (Endsley, 2021, p. 434). One may be concerned that if a user is attending to the virtual content, the physical environment may be ignored. Paying attention to some aspects of the environment can cause users to stop scanning other areas of their environment. This is known as attentional tunneling and can undermine situational awareness (Endsley, 2021). However, if AR and MR is developed correctly, this can be avoided (e.g., important elements of the real environment are not obstructed by virtual content, it is intuitive to the point that the cognitive load on the user is lessened, and if virtual content is appropriately integrated into the environment). Literature has shown that effective AR and MR applications can aid in situation awareness (Stanney et al., 2021). Some examples of this are included in the security domain (Lukosch et al., 2015), UAV operators (Ruano et al., 2017), drivers of autonomous vehicles (Lindemann et al., 2018), and the completion of a safety-critical system, marine transportation (Rowen et al., 2019).

Cybersickness

Cybersickness is, "sensations of nausea, oculomotor disturbances, disorientation, and other adverse effects associated with virtual environment exposure" (Stanney et al., 2014). Cybersickness, or as it is sometimes described as simulator sickness, as a result of virtual reality experiences has been extensively researched and has been discovered to be quite common and side effects can range from mild to severe (Kennedy et al., 1993;

Dużmańska et al., 2018; Stanney et al., 2014; Stanney et al., 2021). One of the most popular theories of why people experience this is sensory conflict theory. This theory explains that people may feel symptoms as a result of receiving conflicting information from different senses (e.g., driving or running in a VR environment will send information through the visual system that movement is happening, but if the user is not physically moving in real time, their kinesthetic and vestibular senses do not receive this same information). In virtual environments, it has been found that display or technology issues (e.g., tracking errors, lag, and flickers) and individual differences (e.g., age, gender, illness, and position in the simulator (sitting or standing)) can cause simulator sickness (LaViola, 2000). Even though users of AR and MR may experience these issues or individual differences, cybersickness has been found to be less severe for these devices when compared to VR (Vovk et al., 2018). This could possibly be due to the real-world environmental cues available to users with AR and MR. They are able to walk around their environment more freely than in VR and have cues from a real horizontal line and physical objects for their visual systems to reference. It is less likely that their different sensory systems will be in conflict. Cybersickness is a physiological symptom that may affect some users, but likely only to a very mild extent.

Privacy

As AR and MR technologies are being developed, privacy has become more of a core issue. These devices function by scanning the user's environment, sometimes constantly in the case of HMDs. These spatial mappings can create a digital twin, an exact virtual replica of the user's space and upload it to the AR cloud. Other personal information can be collected. For example, educational applications like Insight Heart can collect biometric data from a user's smartwatch (Anima Res, n.d.) and location-based application such as Pokémon Go can request to view the user's geolocation from their mobile device (Niantic, 2022). It is important for a user to be aware of how information about their home, workplace, etc. is collected, stored, and protected in order to trust this technology. Some companies, such as Microsoft, Meta, and Google have addressed these issues with privacy guidelines (Hillmann, 2021).

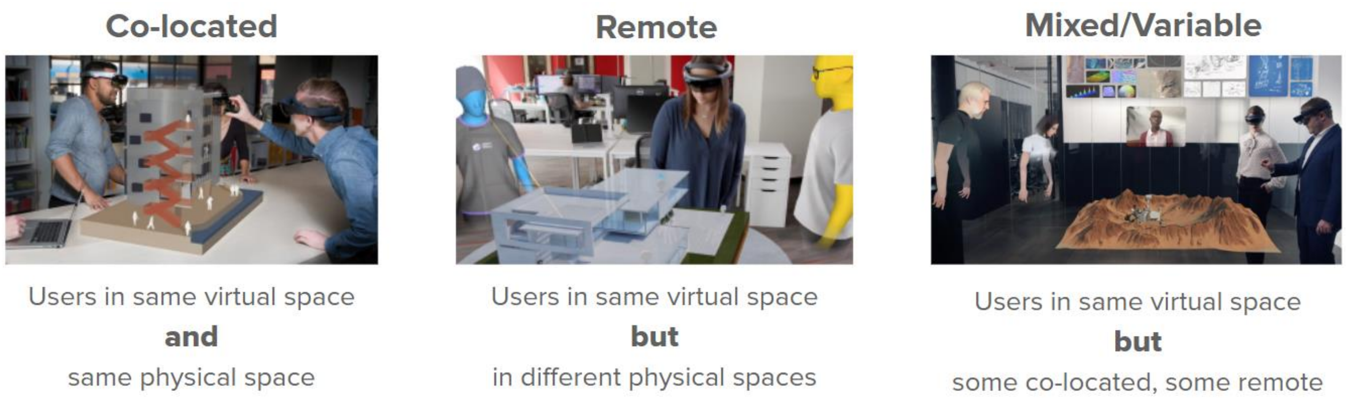
Other privacy concerns may appear as multi-user applications become more populated. A study that asked HoloLens users to brainstorm potential privacy concerns in a multi-user application found that users were concerned by what other AR users could see, interactions AR users could have on non-AR users around them and feeling powerless over their ownership of virtual objects and spaces (Lebeck et al., 2018).

Collaboration

As AR and MR applications are being developed, more multi-user applications are being created. A systematic review of collaborative MR technology between 2013-2018 showed an increase in the number of papers written about collaborative MR applications (de Belen et al., 2018). The largest types of collaborations setup included remote (users located in the same virtual space, but different physical spaces), collocated (all the users located in the same virtual and physical space), and variable (users located in the same virtual space but some may be collocated and others remote) (de Belen et al., 2018). Each of these setups are visualized in Figure 10.

Figure 10

Visualizations of Different AR and MR Collaborative Set-Ups



Note. Images from Geospatial World News Desk (2016, November 8). *Trimble SketchUp Viewer for Microsoft HoloLens enables users to experience designs.* https://www.geospatialworld.net/news/trimbles-sketchup-viewer-microsoft-hololens-enables-users-experience-designs/, Lee, N. (2019, August 8). *Spatial's collaborative AR platform is basically FaceTime in 3D. https://www.engadget.com/2018-10-24-spatial-augmented-reality-3d.html*, and Microsoft. (n.d.) *Get Prism by Object Theory.* https://askus.baker.edu/faq/218130#:~:text=Company%20Name.,Retrieved%20from%20website%20address.

It is not necessary for all users to be using the same device; some users may join using a VR display, MR HMD, AR mobile device, desktop, etc. to view the information. Users may even interact synchronously or asynchronously depending on how the application allows users to interact. There are a few considerations to this environment that could affect users' experiences, such as how users interact with one another and communicate, how device type plays a role in the experience, and the amount of ownership of virtual content is necessary for each user (e.g., should everyone be able to alter all of the virtual content seen, or should it be specified by person?). Literature suggests that visual content is crucial for collaborative tasks and improves the usability and feeling of co-presence within the collaborative application (Aschenbrenner et al., 2019; Kim et al., 2019); AR-to-AR and AR-to-VR collaboration tends to show higher degrees of presence, collaboration, and embodiment than AR-to-Desktop collaboration (Pan et al., 2018); shared virtual content promotes a common ground for users to discuss and avoid miscommunication when completing tasks (Müller et al., 2017).

Inclusive Design

This area is important to consider with any device, but especially AR and MR as these are becoming training, rehabilitation, and environmental enhancement tools for older adults, those with intellectual disabilities, and hospital patients (Conner et al., 2020; Derby & Chaparro, 2020; Hillmann, 2021). There is always a sociological impact with design as well. Hardware may be unintendedly exclusive if it is not customizable to fit different head or face shapes, are incapable with perceptual and safety aids (e.g., glasses, safety equipment, and hearing aids, etc.). By ensuring that inclusion, diversity, and accessibility is supported with AR and MR devices and application, it can make it more accessible to a wider population and better their experiences with these devices.

Evaluating Usability with Heuristics

A heuristic evaluation is a method used to assess the usability of a product. It involves the use of a set of usability heuristics that evaluator(s) can follow to assess aspects of the interface. These heuristics are a set of best practices that are often seen as rules of thumb (Quiñones et al., 2018; Nielsen & Molich, 1990). Two of the most common heuristic evaluations are Nielsen's 10 Heuristics and Schneiderman's Eight Golden Rules (Nielsen, 1994a; Schneiderman, 1998). Nielsen's 10 Heuristics asks the evaluator to rate the system based on 10 categories: visibility of system status, match between system and the real world, user control and freedom, consistency and standards, error prevention, recognition rather than recall, flexibility and efficiency of use, aesthetic and minimalist design, help users recognize, diagnose, and recover from errors, and help and documentation (Nielsen, 1994a). Schneiderman also gives the rater categories to base their evaluation on, but simplifies it down to just eight: Strive for consistency, enable frequent users to use shortcuts, offer informative feedback, design dialogue to yield closure, offer simple error handling, permit easy reversal of actions, support internal locus of control, and reduce short-term memory load (Schneiderman, 1998). These heuristics cover a lot of useful usability issues such as feedback, memory issues, and error handling and prevention, however, there are additional aspects that can alter the users' experience when using specific technologies. Looking specifically at AR and MR for an example, physical comfort and safety are not addressed by these heuristics.

Because of this issue, professionals have developed usability heuristics and checklists to be more specific. These have varied across nearly every domain. There are heuristics that focus on medical devices (Aabel & Abeywarna, 2018), mobile AR (Ko, Chang, Ji, 2013), speech-based smart devices (Wei & Landay, 2018), E-Banking Websites (Baños Díaz & Zapata Del Río, 2018), older-adult specific devices (Petrovčič, Taipale, Rogelj & Dolničar, 2017) and many more. These have also ranged from a short generic heuristics list like Nielsen's and Schneiderman's to specific and comprehensive checklists that can be upwards to 300 items. Heuristics and guidelines have been developed specifically for AR and MR technology.

Augmented Reality Heuristics

Previous work had been conducted to complete a literature review of current AR and MR heuristics (Derby & Chaparro, 2022). This work showed that heuristics and guidelines have been developed to be either broad enough to cover AR applications in general (Kalalahti, 2015; de Paiva Guimarães & Martins., 2014) or defined for specific use cases such as mobile AR games (Endsley et al., 2017). Organizations who are creating this technology have also developed their own, publicly accessible, design guidelines. These companies include Apple, Microsoft, and Magic Leap (Apple, 2021; Magic Leap, 2023; Microsoft, 2022, June 7). AR heuristics and guidelines that were not found in the original literature search included those for instruction-based tasks (Yim & Seong, 2010; Rolim et al., 2015), mobile tourism applications (Shukri et al., 2017), AR user interfaces (Blokša, 2017), AR user interactions (Ortman et al., 2012), contrast ratios (Koreng, & Krömker, 2019), manual assembly tasks (Jeffri & Rambli, 2020), contextual learning, learning support, and education (Santos et al., 2015; Fallahkhair & Brito, 2019; Al-Obaidi & Prince, 2022), kindergarten applications (Tuli & Mantri, 2020), tabletop AR games (Nilsen et al., 2006), and AR video games in smartphones (Chang et al., 2019). These heuristics are updating as the technology improves. For example, in 2011, Wetzel et al.'s design guidelines for Augmented Reality Games were updated from the 2008 design guidelines to include heuristics regarding five categories (general, virtual element, real world element, social element, and technology and usability guidelines) and changed from 12 to 22 guidelines (Wetzel et al., 2011; Wetzel et al., 2008). Chang et al. (2019) is the only one of these additional heuristics that has been validated using the Quiñones et al. methodology (2018).

Mixed Reality Heuristics

There are fewer published MR heuristics than AR heuristics. A previous literature search found fifteen heuristics and guidelines for AR, but only four that focused on MR technology (Derby & Chaparro, 2022). Other MR heuristics and guidelines include those for XR applications, including VR (Vi et al., 2019), MR and Internet of Things interfaces for the elderly (de Belen & Bednarz, 2019), MR robotic games (Pratticó & Lamberti, 2020). Vi et al.'s UX guidelines for designing HMD XR applications followed an adapted approach of Quiñones et al.'s methodology, steps 1-4 were completed but the steps regarding validation were never discussed (Vi et al., 2019; Quiñones et al., 2018).

Literature Review Summary

This extended literature review clarified the definition that are used to describe AR and MR in this document, "eXtended Reality [XR] solutions are computer-generated immersive environments that provide a spectrum of experiences including... Augmented Reality (AR), which overlays virtualized content onto the real world; Mixed Reality (MR), which in addition to augmenting the real world with virtualized content, allows the virtual content to be aware of and interact with the real world; or vice versa with real objects in a virtual world" (Stanney et al., 2021, p.783). It also included definitions of usability and UX as well as the contributing factors that may affect the UX of AR and MR devices. These may include perceptual factors such as, the effectiveness of provided binocular and monocular cues like occlusion and vergence-convergence; visual acuity that may be affected by size, resolution, and contrast; auditory cues such as spatial audio and sound effects; and tactile cues like vibration feedback and touch. It may also be affected by aspects of the device or application such as intuitive user interactions and feedback; clear and easily accessed help and documentation; accurate integration of real and virtual words by following user expectations; consistency and standards between applications to reduce cognitive load and increase learnability of the application; physical comfort and safety of the user; privacy aspects that address users' concerns; and promoting inclusion, diversity, and accessibility through design.

Chapter 3

The Current Study

Purpose

There is a gap in the literature for comprehensive and validated usability heuristics that can be used to assess the usability of any AR and MR application and/or device. The purpose of the current study was to create and validate a new usability heuristic checklist for AR and MR device and application evaluation.

Research Questions

This current study used an eight-step methodology to develop and validate usability/user experience heuristics (Quiñones et al., 2018). This process aimed to answer the following research questions:

1. Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more usability/UX issues** than the use of a control heuristic checklist?
2. Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more issues that qualify as severe/critical** than the use of a control heuristic checklist?
3. Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more domain specific issues** than the use of a control heuristic checklist?
4. Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **similar usability/UX issues as a usability study** of the same application?

Research Design

Quiñones et al. established a methodology to create and validate usability/UX heuristics (2018). This methodology involves eight steps as shown in Table 1 (Quiñones et al., 2018). These steps involve an in-depth look at the current literature surrounding the topic, identifying common UX and usability issues within this domain, mapping those issues with heuristic items, and validating the resulting heuristic checklist through expert judgements, heuristic evaluations, and user testing. Quiñones et al. recommend at least two iterations to refine and improve a set of heuristics twice with different case studies (2019). These iterated steps are less prescriptive to give researchers flexibility based on resources and necessity. For example, the main activity in step 8 (refinement) is to revise the heuristic checklist based on the findings from the validation stage. If the heuristic checklist scored poorly during the validation stage and many changes are required, the authors recommend that validation should be completed again in an iterative manner (Quiñones et al., 2019). However, it is only necessary for one of the three validation methods in step 7 to be completed again as opposed to completing all three validation methods again (expert judgements, heuristic evaluations, and user testing) (Quiñones et al., 2019).

Table 1

Steps in the Eight-Step Methodology to Validate Usability/UX Heuristic Checklists

Step	Brief Description
1. Exploratory	Literature Review about the domain, features, UX attributes, and existing heuristics
2. Experimental	Analyze Literature, find what's missing
3. Descriptive	Select & prioritize important topics of information
4. Correlational	Match features of domain with UX attributes & existing heuristics
5. Selection	Keep, adapt, add, and/or discard heuristics
6. Specification	Formally specify set of new heuristics
7. Validation	Validate effectiveness & efficiency through: 1) Expert Judgments 2) Heuristic Evaluations 3) User Testing
8. Refinement	Refine & improve new set of heuristics by repeating steps

Note. Altered from Quiñones, D., Rusu, C., & Rusu, V. (2018). A methodology to develop usability/user experience heuristics. *Computer Standards & Interfaces, 59,* 109-129. https://doi.org/10.1016/j.csi.2018.03.002

Derby & Chaparro (2022) developed an AR and MR usability heuristic checklist following the eight-step methodology (Quiñones et al., 2018). The checklist was validated with two applications on HMDs (the Magic Leap 1 and HoloLens 1) and a mobile phone. This validation yielded 11 total heuristics and 94 items. Appendix A lists the heuristics and checklist items.

Hardware and Applications Used for the Current Study

This current work extends the effort by Derby & Chaparro (2022). During the validation process, it was discovered that this heuristic checklist could be broadened to ensure it accurately represents a wider range of AR and MR technologies and use cases. AR and MR usability aspects that were covered in the Derby & Chaparro heuristic checklist include: help and documentation, cognitive overload, consistency and standards, comfort, feedback, user interaction, and recognition rather than recall. This heuristic checklist included items and adequately validated these usability aspects through the validation stage. Usability aspects that were somewhat covered include: unboxing and set-up, integration of physical and virtual worlds, device maintainability, privacy, and different forms of sensory output (e.g., spatial audio and tactile feedback). Some of these usability aspects had items included in the heuristic checklist but could have been validated with different applications or devices to ensure the validity of the items. For example, this heuristic checklist was tested using mobile phones and HMDs, but other technologies with different set-up processes like smart glasses and head-mounted VST displays were not used. Additionally, some usability issues that occur when integrating real and virtual worlds were identified during testing (e.g., if real-world elements are easily distinguishable from virtual elements), but no applications were used that required the user to place virtual elements onto real world objects. Instead, the virtual elements in these applications were just floating in the real space. More items may be necessary to add to the checklist to ensure that usability issues are adequately addressed for these types of applications. Usability aspects that were not adequately covered in the Derby & Chaparro checklist include: collaboration, safety, and inclusive design. These usability aspects were not evaluated during the validation phase due to the lack of relevant features that the chosen applications included. Additionally, no items were included in the checklist that addressed usability aspects related to inclusive design. Table 2 describes the AR and MR usability issues that were used in the validation of the first version of the heuristic checklist and those that needed to be further evaluated.

Table 2

Usability Aspects for AR and MR Mapped to the Derby and Chaparro (2022) AR and MR Heuristic Checklist

AR & MR Usability Aspect	Was it Covered & Validated in Derby & Chaparro, 2022?
Unboxing/Set-Up	Somewhat. Apps were tested on HMD & mobile phones, but have not specifically tested to see if this is valid for just a headset evaluation (e.g., unboxing the Microsoft HoloLens) or with other technology such as smart glasses.
Help & Documentation	Yes
Cognitive Overload	Yes
Integration of Physical & Virtual Worlds	Somewhat. Items related to overlaying virtual objects on real world objects and interacting with both at the same time are incorporated, but have not been specifically tested to see if those items are valid with applications that incorporate such tasks.
Consistency & Standards	Yes
Collaboration	No
Comfort	Yes
Feedback	Yes
User Interaction	Yes
Recognition Rather than Recall	Yes
Device Maintainability	Somewhat. Apps were tested on HMD & mobile phones, but have not specifically tested to see if this is valid for just a headset evaluation (e.g., unboxing the Microsoft HoloLens).

AR & MR Usability Aspect	Was it Covered & Validated in Derby & Chaparro, 2022?
Privacy	Somewhat. Items related to collaboration and privacy from other users are included, but no items regarding data privacy and how data is stored are included. This is especially important for use cases that include digital twins and spatial mapping data.
Safety	No
Sensory Output (Tactile, Audio)	Somewhat. Basic audio aspects are covered "is the volume adjustable" "are auditory features understandable" "are captions available" under consistency & standards. However, spatial audio and use of audio as feedback is not covered. Tactile feedback is not currently covered by the checklist.
Inclusive Design	No

To ensure that the validation process in this current work included the consideration of usability aspects missing from the Derby & Chaparro (2022) heuristic checklist, the chosen applications and devices were mapped to the missing usability aspects. The chosen applications and devices included the following: the Meta Quest Pro (HMD) and ShapesXR, a collaborative prototyping application; the Magic Leap 1 (HMD) and Wayfair Spaces, a retail application that allows users to see virtual furniture in their own space; Magic Leap 2 (HMD) and Fan Blade Replacement, a training application for aircraft maintenance; mobile phone and Google Maps Live View, an AR navigation application; and Epson Moverio BT-300 (smart glasses) and a variety of applications on the device. These chosen devices and applications included a wide variety of interaction methods, use cases, and features that adequately covered the areas that were either not or somewhat covered in the Derby & Chaparro heuristic checklist (2022). The applications and devices used in this current study and how they map to the usability aspects are depicted in Table 3, and the hardware's technical specifications are shown in Table 4.

Table 3

Applications and Devices Mapped to Usability Aspects in the Derby and Chaparro (2022) AR and MR Heuristic Checklist

	Epson Moverio BT-300 (Smart Glasses)	**Magic Leap 1 (HMD)**	**Magic Leap 2 (HMD)**	**Meta Quest Pro (HMD)**	**iPhone 13 Pro or Android Galaxy S8 (Mobile Phone)**
Device Type	AR	MR	MR	XR	AR
Application	Variety of Apps	Wayfair Spaces	Fan Blade Replacement	ShapesXR	Google Maps Live View
Application Type	Marker-less	Marker-less	Marker-based	Marker-less	Location-based
Degree Of Movement	Minimal (standing is useful but not required)	Moderate (some light walking)	Moderate (some light walking)	Moderate (some light walking)	High (heavy walking)
Input(s) Used	Head tracking Controller input	Controller input	Controller input Gesture input Voice input	Head tracking Eye tracking Facial tracking Controller input Voice input	On-screen tap
Unboxing and Set-up	Yes	Yes	Yes	Yes	Yes
Integration Of Physical & Virtual Worlds	No	Yes	Yes	No	Yes
Collaboration	No	No	No	Yes	No
Device Maintainability	Yes	Yes	Yes	Yes	Yes
Privacy	No	Yes	No	Yes	Yes

	Epson Moverio BT-300 (Smart Glasses)	Magic Leap 1 (HMD)	Magic Leap 2 (HMD)	Meta Quest Pro (HMD)	iPhone 13 Pro or Android Galaxy S8 (Mobile Phone)
Safety	Yes	Yes	Yes	Yes	Yes
Inclusive Design	Yes	Yes	Yes	Yes	Yes
Additional Sensory Features (Tactile, 360-audio)	No	No	No	Yes	Yes

Note. A "Yes" means this usability aspect is either a feature or a consideration for the device. A "No" means it is not a feature or a consideration for this use case.

Table 4

Technical Specifications of the Five Devices Used in the Current Study

Device	Epson Moverio BT-300	Magic Leap 1	Magic Leap 2	Meta Quest Pro	iPhone 13 Pro or Android Galaxy S8 (Mobile Phone)
Type of AR or MR	HMD Optic see-through	HMD Optic see-through	HMD Optic see-through	HMD Video see-through	Mobile Device Video see-through
Tethered or Untethered	Tethered	Tethered	Tethered	Untethered	N/A
Input Methods Used for App	Controller, head-tracking	Controller	Controller, voice control, hand gestures	Controller, face tracking	Tap on screen
Refresh Rate	30Hz	122Hz	120Hz	90Hz	10-120Hz
Display Resolution	921,000 pixels (1280 RGB x 720)	1280 x 960 pixels per RGB eye	1440 x 1760 pixels	1800 x 1920 pixels per eye	iOS: 2532 x 1170 pixels Android: 2960 x 1140 pixels
Field of View	23°D	40°H x 30°V x 50°D	44.6°H x 53.6°V x 70°D	106°H x 96°V	N/A

Note: The data for Epson Moverio BT-300 is from Epson. (2023). *Epson Moverio BT-300 Smart Glasses*. https://www.epson.com.au/products/ProjectorAccessories/Moverio_BT-300_Specs.asp. The data for Magic Leap 1 is from Leon. (2022, May 21). *Microsoft HoloLens 2 vs Magic Leap One: a comprehensive comparison.* https://vrx.vr-expert.com/microsoft-hololens-2-vs-magic-leap-one-a-comprehensive-comparison/. The data for Magic Leap 2 is from Magic Leap. (n.d.a.). *Magic Leap 2.* https://www.magicleap.com/magic-leap-2. The data for Meta Quest Pro is from Meta. (n.d.). *Meta Quest Pro Tech Specs.* https://www.meta.com/quest/quest-pro/tech-specs/. The data for the mobile phones is from Apple. (2023b, May 10). *iPhone 13 Pro – Technical Specifications. https://support.apple.com/kb/SP852?locale=en_US* and Tom's Guide Staff. (2018, October 3). *Samsung Galaxy S8 user guide: tips, tricks, and how-tos.* https://www.tomsguide.com/us/samsung-galaxy-s8-guide,review-4330-8.html.

The rest of this chapter describes the method and results found during each of the 8 steps completed in the current study.

Step 1 – Exploratory Stage and Step 2 – Experimental Stage

Method

Steps 1 and 2 involved a review of the current literature, with a focus on current heuristics, guidelines, definitions, and results from usability studies and experiments within the domain. This is done to compile definitions and features about the domain, current heuristics, and usability and user experience (UX) principles related to AR and MR applications and devices.

The following are the search terms that were used to expand on Derby & Chaparro heuristic checklist (2022). A search for current AR and MR developer design guidelines for user experience, inclusive design, and user safety was also conducted.

- Heuristics AND (AR OR MR)
- Privacy AND (AR OR MR)
- Privacy AND (AR OR MR) AND Multi-User
- Collaboration AND (AR OR MR) AND UX
- Cybersickness AND (AR OR MR)
- Motion sickness AND (AR OR MR)

Results

As a result of step 1, 10 heuristics, design guidelines, and standards were found and deemed useful for the development of the heuristic checklist. These focused on areas including: XR HMD applications, AR safety, virtual environments (VEs), XR accessibility, mobile AR, and XR ethics. These articles are shown in Table 5. As a result of step 2, 42 user studies, experiments, and review articles were found that were useful for defining features about the domain and usability and UX principles related to AR and MR applications and devices. Some of the areas that these articles focused on included: AR/MR user privacy, AR/MR/XR collaborative apps, AR/MR physiological risks such as cybersickness, MR ethics, and XR user experience. These articles are summarized in Appendix B. These articles were found in addition to those found in Derby & Chaparro's heuristic checklist (2022).

Table 5

Heuristics, Guidelines, and Standards Found as a Result of Step 1 – Exploratory Stage

Citation	Area of Focus	Type of Article
Vi et al., 2019	XR HMD applications	Heuristics
AREA, 2021	AR Safety Best Practices	Guidelines
Stanney et al., 2003	Virtual Environments (VE)	Heuristics
Stanney, 2021	VE Usage Protocols	Guidelines
O'Connor et al., 2021	XR Accessibility User Requirements	Standards
Magic Leap, 2018	AR and MR Accessibility	Developer Kit (Guidelines)
Meta, 2019a	Best Design Practices for Mobile AR	Developer Kit (Guidelines)
Meta, 2019b	Mobile AR	Developer Kit (Guidelines)
Apple, 2022	Accessibility	Developer Kit (Guidelines)
McGill, 2021	XR Ethics	Standards

Step 3 – Descriptive Stage

Method

In step 3, the information gained from the literature search was organized. Important information from the domain, heuristics, and usability and user experience (UX) principles are selected and prioritized. Information was grouped by:

- Information about AR and MR (definitions, devices, etc.)

- Features specific to AR and MR
- Usability/UX Attributes
- Existing sets of heuristics
- Known usability problems

Information was prioritized by using the following scale: (3) highly important, (2) somewhat important, (1) not important. After this was created, it was compared to current heuristic lists in the domain.

Results

It was determined that current heuristic lists do not sufficiently evaluate the AR and MR domain. The output of this step is in Appendix C and an example excerpt is shown in Table 6. In this domain, many checklists are very specific (e.g., only apply to mobile AR games). These checklists would not cover a similar domain area as our goal (any visual AR or MR application and device). As a result, the heuristic checklist that was most comparable was used as a control checklist to compare to the experimental checklist for the new checklist development (de Paiva Guimarães & Martins, 2014). This heuristic checklist has 12 heuristics and 21 checklist items. They can be looked at in detail in Appendix D.

Table 6

Excerpt of the Results from Step 3 – Descriptive Stage

Topic	Collected Information	Selected Information	Priority
Features specific to AR & MR	General features of AR/MR: User-Information (enjoyment, visibility, etc.); User-Cognitive (consistency, learnability, predictability, etc.); User-Support (error management, help & documentation, etc.); User-Interaction (feedback, responsiveness, low physical effort, etc.); User-Usage (navigation, availability, exiting, etc.) (Ko et al., 2013) Specific features of AR/MR: Content accuracy (info about system AND/OR lag/jitter etc.); privacy of content (for collaborative spaces); virtual/environmental distinctions; energy usage/battery life; safety (cybersickness, eye strain, heat issues, safety of accessories like tethers, etc.); privacy of user data; situational awareness	Both general and specific features will be kept and used for the heuristic checklist.	(3) Highly important

Step 4 – Correlation Stage

Method

In step 4, by using the information that was collected in step 3, features of AR apps and devices were matched to existing heuristics, and usability and user experience (UX) principles. For each feature, an experimental heuristic was created to evaluate it. The first version of the heuristic checklist was used to map to these features and principles.

Results

The AR and MR features and usability/UX attributes were mapped to the existing heuristic categories from the first version of the heuristic checklist. This is depicted in Table 7. All of the heuristics could be matched to a feature and usability/UX attribute, so no additional heuristics were created during this step.

Table 7

AR and MR Features Mapped to Usability and UX Attributes and Heuristics from Derby & Chaparro (2022)

General Feature	Feature Details	Usability/UX Attribute	Control Heuristics (de Paiva Guimarães & Martins, 2014)	Derby & Chaparro, 2022 Heuristic Checklist
User-Information	Defaults; Enjoyment; Familiarity; Hierarchy; Multi-modality; Visibility	satisfaction; desirable; Simplicity (UX); efficiency (UX)	1. Visibility of System Status 3. Satisfaction 4. Aesthetic and Minimalist Design	ARH1 Set-Up; ARH3 Aesthetic & Minimalist design; ARH8 Visibility of System Status
User-Cognitive	Consistency; Learnability; Predictability; Recognition	efficiency; learnability; memorability; Usable (UX); findable (UX); Useful (UX); Consistency (UX); workload reduction (UX)	4. Aesthetic and Minimalist Design 5. Match Between System & Real World 6. Consistency & Standards 8. Recognition Rather than Recall 12. Accuracy	ARH3 Aesthetic & Minimalist design; ARH5 Consistency & Standards; ARH10 Recognition rather than recall
User-Support	Error Management; Help & documentation; Personalization; User Control	effectiveness; satisfaction; errors; Usefulness (UX); error prevention & handling (UX)	2. User Control & Freedom 7. Error Prevention 10. Help Users Recognize, Diagnose, and Recover from Errors	ARH2 Help & Documentation; ARH9 User Interaction & Flexibility

General Feature	Feature Details	Usability/UX Attribute	Control Heuristics (de Paiva Guimarães & Martins, 2014)	Derby & Chaparro, 2022 Heuristic Checklist
User-Interaction	Direct manipulation; Feedback; Low Physical Effort; Responsiveness	satisfaction; efficiency; Usable (UX); Communication (UX); accessible (UX)	1. Visibility of System Status 9. Flexibility & Efficiency of Use	ARH7 Comfort; ARH8 Visibility of System Status; ARH9 User Interaction & Flexibility; ARH4 Integration of Physical & Virtual Worlds; ARH6 Collaboration
User-Usage	Availability; Context-Based; Exiting; Navigation	efficiency; satisfaction; Usable (UX)	11. Environment Configuration	ARH5 Consistency & Standards; ARH6 Collaboration; ARH11 Device maintainability

Step 5 - Selection Stage

Method

Step 5 was completed to refine the experimental heuristic checklist. Heuristics and checklist items that were compiled in step 4 were adapted, added, or discarded. Since this work was a continuation of a previously developed AR and MR heuristic checklist (Derby & Chaparro, 2022), these heuristics and checklist items largely contributed to this step. These checklist items were used as a starting point for the new experimental heuristic checklist.

Results

A total of 13 checklist items were added (13 examples were added along with these items) and 34 item examples were rephrased from the Derby & Chaparro (2022) heuristic checklist. These changes were based on the literature review conducted in steps 1 and 2 on the usability aspects that were not originally included in the heuristic checklist (unboxing and set-up, integration of physical & virtual worlds, collaboration, device maintainability, privacy, safety, inclusive design, and additional sensory features (tactile, 360-audio)). These changes are shown in Tables 8 and 9. Two items were eliminated (both located in the "comfort" heuristic). These items were deemed similar to each other, so they were combined into one new item. A total of 11 heuristics and 105 checklist items were created for the experimental heuristic checklist.

Table 8

Number of Items Added to the Heuristic Checklist in Step 5 – Selection Stage

Area of interest:	**Collaboration**	**Privacy**	**Accessibility & Inclusion**	**Safety**	**Cybersickness**	**Total Added by Heuristic**
Heuristic: Collaboration	4	2	2	0	0	**8**
Heuristic: Integration of Physical & Virtual Worlds	0	1	1	0	0	**2**
Heuristic: Comfort	0	0	0	2	0	**2**
Heuristic: Consistency & Standards	0	0	1	0	0	**1**
Total Added by Area of Interest	**4**	**3**	**4**	**2**	**0**	**Total Additions: 13**

Table 9

Number of Rephrased Examples in the Experimental Heuristic Checklist

Area of Interest:	Collaboration	Privacy	Accessibility & Inclusion	Safety	Cybersickness	Total by Heuristic
Heuristic: Collaboration	0	6	0	0	0	**6**
Heuristic: Integration of Physical & Virtual Worlds	1	0	0	1	0	**2**
Heuristic: Comfort	0	0	5	6	3	**14**
Heuristic: Consistency & Standards	1	0	2	1	0	**4**
Heuristic: User Interaction	0	0	2	2	0	**4**
Heuristic: Help & Documentation	0	0	1	1	0	**2**
Heuristic: Cognitive Overload	0	0	1	0	0	**1**
Heuristic: Unboxing & Set-Up	0	0	0	1	0	**1**
Total by Area of Interest	**2**	**6**	**11**	**12**	**3**	**Total: 34**

Step 6 – Specification Stage

Method

Step 6 incorporates the information gathered from steps 3–5. Formalized heuristics are defined, and the following is created for each heuristic in the experimental heuristic checklist:

- Id: Heuristic's identifier.
- Priority: Value that identifies how important the heuristic is in the evaluation of a specific aspect or feature. The value can be (1) Useful: Heuristic further improves the usability/UX; (2) Important: Heuristic evaluates a relevant aspect; or (3) Critical: Heuristic evaluates a crucial aspect.
- Name: Heuristic's name.
- Definition: A brief but concise definition of the heuristic.
- Explanation: Detailed explanation of the heuristic.
- Application feature: Feature or aspect of the specific application domain that is evaluated with the heuristic.
- Examples: Examples of violation of and compliance with the heuristic. Include an image that graphically explains the problem.
- Benefits: Expected usability/UX benefits when the heuristic is satisfied.
- Problems: Anticipated problems of heuristic misunderstanding.
- Checklist: Items or criteria that are associated with the heuristic.
- Usability/UX attribute: Usability/UX attribute that is evaluated with the heuristic.

- Heuristics related: Set (or sets) of heuristics on which the heuristic is based, along with the authors and the references.

Clearly defined heuristics are created through the completion of this step.

Results

Eleven heuristics were formally defined as a result of this step. A sample result of this step is shown in Table 10 and the full list of definitions can be found in Appendix E.

Table 10

Sample of a Formal Definition of a Heuristic

ID	**H1**
Priority	(2) Important
Name	Unboxing & Set-Up
Definition	Getting started with the AR/MR device/application should be easy to identify, complete, and a positive experience.
Explanation	A user's first experience is very important because it sets the tone and future experience with the device/application. If the set-up process is difficult, it could deter the user from any future use with the device.
Application feature	User-Information; enjoyment; familiarity; learnability
Examples	When the user takes the device out of its packaging, it should be clear what each of the pieces do and how to assemble them (if necessary). Unboxing the device should be easy to do, the user should not have to struggle with taking the device out of its packaging. For a mobile experience, it should be clear how to interact with the AR/MR content (e.g., a call to action like a QR code or other identifying information).
Benefits	An easy unboxing experience can create a seamless and positive first experience. This can directly affect the user's first impressions with the device, setting a positive tone for future interactions.
Problems	Misunderstanding of this heuristic can cause the user to feel confused or frustrated and can lead to abandonment of use. If this unboxing and set-up process is difficult, there is potential that the device could be accidentally damaged.

Checklist Items	Element	Checklist item	More information and Examples
	Device Unboxing (first-time usage)	Is the unboxing process a positive experience?	This can be both physical and emotional. The physical unboxing of the device itself is easy, understandable, and/or does not harm the device. Emotionally, unboxing is exciting or enjoyable for the user.
	Device Set-Up & Configuration (first-time usage)	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?	This can be in the form of a welcome page, initial introduction, tutorial, etc.
		Is a quick start guide available with the device?	A quick start guide is a brief instruction manual that describes how to begin using the AR device and risks to keep in mind when using the device (e.g., be aware of your environment, take breaks, etc.). This is often provided in the original box with the device, but can also be provided online. It is often more brief than a tutorial, only providing a simplified overview of how to start using the device.

	Application Set-Up	Is a call to action (QR code, instructions to use AR, etc.) clearly marked in the physical space?	This is often done for virtual elements that are pinned to a specific location (E.g., If the only way to see an AR message is to take a picture of a real-world sign with the app). This often applies to marker-based applications, and may not apply to all applications.
Usability attributes	Satisfaction; desirable		
UX factors	Simplicity; efficiency		
Set of heuristics related	Santos, 2016; de Paiva Guimarães & Martins, 2014		

Overall Discussion of Steps 1-6

Steps 1 through 6 were completed to create and formally define the experimental heuristic checklist. Step 1 and step 2 were completed to gain an overview of the domain, heuristics that currently exist within it, and usability and UX principles relevant to the domain. The data collected at this point was the first step to defining new checklist items in the experimental heuristic checklist. Ten heuristic lists and 42 user studies, experiments, and review articles were examined. Themes and findings found from the literature review were integrated into the experimental heuristic checklist items to ensure that the following areas of interest were covered: the merging of real and virtual worlds, collaboration, safety, privacy, and inclusive design.

Step 3 helped identify information about the domain, usability/UX attributes, usability attributes, and relevant AR or MR heuristics. These were crucial to address with the experimental heuristic checklist, as it included the most important factors that impact the usability of AR and MR applications and devices. This step also identified which heuristic checklist to use as a control to compare the experimental checklist with during the validation stage. A heuristic checklist with 12 heuristics and 21 checklist items was chosen as the control heuristic checklist because it was a validated heuristic checklist currently being used in the domain, and it covered the most usability issues that were identified during steps 1 and 2 (de Paiva Guimarães & Martins, 2014). Step 4 took an in-depth look at this heuristic checklist and the heuristic checklist developed previously for AR and MR (Derby & Chaparro, 2022). The control checklist and (Derby & Chaparro, 2022) included checklist items that related to each of the AR and MR general features, so no additional items were added in this step. However, based on the detailed literature review on AR and MR, it was deemed necessary to add items and examples to the heuristic checklist in step 5. A total of 34 edits were made and the experimental heuristic checklist was created with 11 heuristics and 105 items. This experimental heuristic checklist was then validated in step 7.

Step 7 – Validation Stage

Step 7 was completed to validate the experimental heuristic checklist. Three different tests (heuristic evaluation, expert judgement, and user testing) were completed to investigate user perceptions and show that the new heuristics are effective, easy to use, and necessary. According to Quiñones et al., only the heuristic evaluation test is required for validation (2018). The other two validation tests (expert judgement and user testing) are optional but can further refine the heuristic checklist. In this study, all three validation steps were completed because the additional information from expert judgement and user tests were expected to result in impactful changes to the heuristic checklist. The applications discussed at the beginning of this chapter and described previously in Table 3 were used for the heuristic evaluations and user tests (ShapesXR with the Meta Quest Pro, Google Maps Live View with mobile phones, Epson Moverio and its multiple applications, Wayfair Spaces on the Magic Leap 1, and the Fan Blade Replacement application on the Magic Leap 2). Expert judges chose to evaluate any application or device of their choosing since they completed this remotely, at home or at work, across the country. The applications that were used for each of these stages are shown in Table 11.

Table 11

Applications and Devices Used for Each of the Three Validation Tests

Application	Device	Heuristic Evaluation	User Test	Expert Judgement
Variety of Local Apps	Epson Moverio BT-300	X	X	App and device were different for each of the experts
Wayfair Spaces	Magic Leap 1	X	X	
Fan Blade Replacement	Magic Leap 2	X	X	
ShapesXR	Meta Quest Pro	X	X	
Google Maps Live View	Mobile Phone	X	X	

Expert Judgement

Method. Five experts were asked to check the validity of the proposed heuristics by assessing its utility, clarity, ease of use, need for checklist, and comments about each heuristic. These experts were professionals who have worked with AR or MR in the past, either through engineering, design, or user experience research. Experts were asked to complete a heuristic evaluation with an AR or MR application of their choice (the chosen applications are described in Table 9) to guide them through the heuristics, and then asked to rate the following scales developed by Quiñones et al. (2018):

Scales completed for each heuristic:

- **Utility**: How useful is the heuristic? (1 – Not at all useful; 5 – Extremely useful)
- **Clarity**: How clear is the heuristic? (1 – Not at all clear; 5 – Extremely clear)
- **Ease of use:** How easy it was to associate identified problems with the heuristic? (1 – Very difficult; 5 – Very easy)
- **Necessity of an additional checklist:** How necessary is it to complement the heuristic with a checklist? (1 – Not at all necessary; 5 – Extremely necessary)
- **Category Importance:** How important is the heuristic? (1) = Useful: Heuristic further improves the usability/UX; (2) = Important: Heuristic evaluates a relevant aspect; (3) = Critical: Heuristic evaluates a crucial aspect

Scales completed with regards to the heuristic checklist as a whole:

- **Ease/Difficulty:** How easy was it to perform the heuristic evaluation, based on this set of usability/UX heuristics? (1 – Very easy; 5 – Very difficult)
- **Intention:** Would you use the same set of usability/UX heuristics when evaluating similar products in the future? (1 – Extremely unlikely that I would use it; 5 – Extremely likely that I would use it)

- **Completeness:** Do you think the set of usability/UX heuristics covers all usability aspects for this type of software products? (1 – Not at all complete; 5 – Extremely complete)

Each of the expert evaluators were provided with a version of the heuristic checklist with scales listed above. Experts were located across the United States and completed their evaluations remotely, either at their home or place of work. Feedback was provided through an excel file that was sent through email. Because experts only had access to AR and MR devices that were around them, they were given the opportunity to pick an application of their choice to complete the heuristic evaluation and provided scales. These experts provided the purpose of the application they chose and device that they used. These ranged from mobile games to in-house training applications on HMDs. These applications and devices are described in Table 12.

Table 12

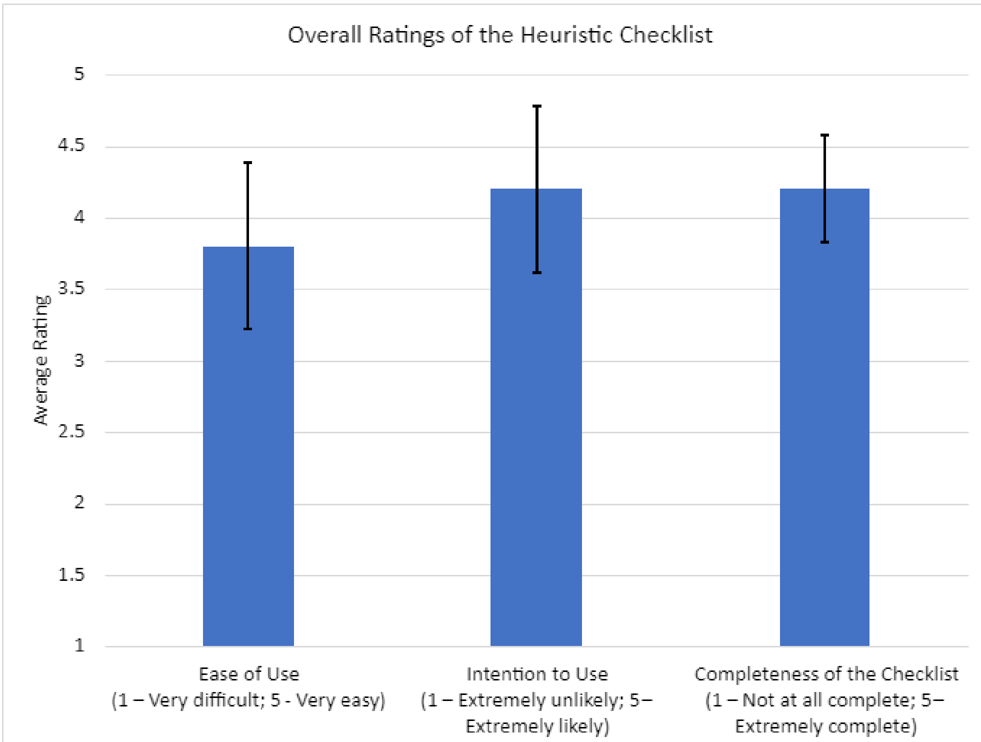
Devices and the Purpose of Applications Used for Expert Judgment Tests

Expert Evaluator	Purpose of Application	Device
1	Educational Games Entertainment	Meta Quest Pro
2	Game	Mobile Phone (Android)
3	Maintenance Training	HoloLens 2
4	Navigation	Mobile Phone (iOS)
5	Training	T-45C MR Simulator

Results. Overall, the expert evaluators rated the heuristic checklist easy to use (*M* = 3.8, *SD* = 1.3*),* mostly complete (*M* = 4.2*, SD* = 1.3), and stated that they were likely to use the heuristic checklist again if they needed to evaluate the usability of AR or MR (*M* = 4.2, *SD* = 0.84). These results are shown in Figure 11. Most heuristics were rated moderately useful, moderately clear, easy to complete, important, and somewhat necessary to include checklist items with the heuristic. A detailed summary of each of the means and standard deviations for each of these scales across heuristics is described in Table 13. Graphs of these results can be found in Appendix F. The most useful heuristics were Comfort, Help & Documentation, and Cognitive Overload. The least useful heuristics were Consistency & Standards, and Device Maintainability. The clearest heuristics were Collaboration, Comfort, and Device Maintainability. The least clear heuristic was Consistency & Standards. The easiest heuristic to evaluate was Device Maintainability and the most difficult was Consistency & Standards. The heuristics that were deemed most necessary to pair with checklist items were Help & Documentation and Comfort. The heuristics that were least necessary to pair with checklist items were Device Maintainability & Consistency & Standards. The most useful heuristics were Unboxing & Set-Up, Cognitive Overload, and User Interaction. The least useful heuristic was Consistency & Standards.

Figure 111

Overall Ratings of the Experimental Heuristic Checklist by Expert Reviewers.



Note: Error bars indicate standard error (SE).

Table 13

Expert Judgement Survey Results

	Utility	Clarity	Ease of Use	Necessity for a Checklist	Category Importance Assigned During Expert Review	Category Importance Assigned During Step 6
Unboxing & Set-Up	4 (1.23)	4.2 (1.3)	3.6 (0.55)	3.2 (1.64)	2.4 (0.55)	2 - Important
Help & Documentation	4.4 (0.55)	4 (0.71)	4 (0.71)	4 (1.73)	2.2 (1.1)	3 - Critical
Cognitive Overload	4.4 (0.89)	4.2 (1.1)	4.2 (0.84)	3.2 (1.64)	2.4 (0.55)	2 - Important
Integration of Physical & Virtual Worlds	4 (1)	4.2 (0.84)	3.8 (0.84)	2.6 (1.34)	2 (0.71)	2 - Important
Consistency & Standards	3 (0)	3 (0)	3 (0)	2 (0)	1 (0)	2 - Important
Collaboration	3.8 (1.3)	4.4 (0.89)	3.6 (1.52)	3.2 (1.64)	1.8 (0.84)	2 - Important
Comfort	4.6 (0.89)	4.4 (0.89)	4.4 (0.89)	3.8 (1.1)	2.2 (0.84)	3 - Critical
Feedback	3.6 (0.89)	3.4 (0.55)	3.8 (0.45)	2.8 (0.84)	1.6 (0.55)	2 - Important
User Interaction	4.2 (0.84)	4.2 (0.84)	4.2 (0.84)	3.2 (0.84)	2.4 (0.55)	3 - Critical
Recognition Rather than Recall	3.4 (0.89)	4.2 (0.84)	4.2 (0.84)	2.2 (1.1)	1.6 (0.55)	2 - Important
Device Maintainability	3 (1.42)	4.4 (0.89)	4.8 (0.45)	1.8 (0.84)	1.8 (0.84)	1 - Useful

Note. The table is formatted as: Mean (Standard Deviation). Items rated on a Likert scale for Utility, Clarity, and Necessity for a Checklist (1 =Not at all [useful, clear, necessary] – 5 = *Extremely* [useful, clear, necessary]). Likert Scale for Ease of use (1 = *Very Difficult* – 5 = *Very Easy*). Importance scale was 3 (Critical), 2 (Important), and 1 (Useful).

Expert evaluators gave feedback about specific checklist items and the heuristic checklist as a whole. This caused changes to the experimental heuristic checklist by rewording items or examples, adding items, and creating a new heuristic called "privacy". For example, in the Unboxing & Set-Up heuristic, an evaluator stated that items should be added for setting up the device and system, "how easy is it to go back and use again?" As a result, an item was added to this heuristic, "Is it easy to set up the device and/or application between uses?" A total of 19 changes were made as a result of this step (4 items added, 4 examples added, 1 item reworded, 4 examples reworded, 1 new heuristic called "privacy" added, 5 items moved to the new heuristic called "privacy" and moving the heuristic tabs closer to each other based on relatedness). A summary of experts' feedback is provided in Table 14.

Table 14

Changes Made to the Experimental Heuristic Checklist from the Feedback Given by Expert Reviewers

Heuristic	Original Checklist Item/Example	Feedback	Change(s) made	Updated Item/Example
Unboxing & Set-Up	Is a call to action (QR code, instructions to use AR, etc.) clearly marked in the physical space?	Smartphone app was required but unsure how to get it - could add to a heuristic as an item or example	Reword item	Is a call to action (QR code, instructions to use AR, instructions to download an app, etc.) clearly marked in the physical space?
Unboxing & Set-Up	N/A	Add components for set up device/system - how easy is it to go back and use again?	Add item (and example)	Is it easy to set up the device and/or application between uses?

Heuristic	Original Checklist Item/Example	Feedback	Change(s) made	Updated Item/Example
Help & Documentation	Does the device's user interface and/or application avoid irreversible errors? e.g., includes a back button.	Parce out. There are times where my app freezes up to where I have to rely on hardware features to take me back home to reload app	Reword Example	This can apply to either the device's user interface and application interface (e.g., a back button or home button).
Cognitive Overload	For example, an application would NOT be easing a user in if they were told to dodge projectiles as soon as they entered the game.	Add item - introduction of more complex/niche features over time (maybe add with "eased into the virtual environment" as an example?)	Reword example	Overwhelming users during their first use of the application should be avoided. Basic interactions should be focused on while more complex features should be integrated over time. An application would NOT be easing a user in if they were told to dodge a variety of projectiles by quickly using combos as soon as they enter the game.
Cognitive Overload	N/A - full heuristic/multiple items; not a specific item	Move Cog overload near comfort since they are related	Move heuristics based on how related they were to one another	N/A - full heuristic/multiple items; not a specific item
Integration of Physical & Virtual Worlds	Is it clear to both users and bystanders when captures and/or recordings are being taken?	This item doesn't belong here	Move item to newly added "privacy" heuristic	N/A - item was moved to a different heuristic

Heuristic	Original Checklist Item/Example	Feedback	Change(s) made	Updated Item/Example
Integration of Physical & Virtual Worlds	N/A - full heuristic/multiple items; not a specific item	Integration of physical & virtual worlds needs expanding (integration with each other and how user actions can affect either/both)	Add items (and examples)	Item added: Is the visual appearance of the real-world environment sufficient to help the user accomplish required tasks? Item added: Do the virtual elements help the user accomplish the required tasks in a meaningful way?
Privacy	Example: Users' privacy should be protected. This includes their data, personalized settings, application-created content, etc.	Include example about authentication and general usage in secure corporate environments (privacy) - e.g., ability to password protect, have different user accounts, work offline	Reword Example	Terminology was added to the example: "Some examples about how this can be done include: creating the ability to password protect a device or application, create different user accounts, or work offline."
Consistency & Standards	Does the device and/or application avoid jargon?	Add item overall language being clear/understandable	Reword item to be broader	Is the language that is used in the device and/or application easy to understand?
Collaboration	N/A - full heuristic/multiple items; not a specific item	I'm torn on my usefulness ratings. I believe privacy heuristics are significantly more important than the others listed here	Addition new heuristic called "privacy" Relevant items moved from collaboration to this new heuristic, and one item added	N/A - full heuristic/multiple items; not a specific item

Heuristic	Original Checklist Item/Example	Feedback	Change(s) made	Updated Item/Example
User Interaction	The device and/or application should be able to adapt virtual elements in a useful manner. If the virtual content needs to be world-locked, allowing the user to walk around it, the device and/or application should support this. This also applies to content that should be head-locked (follows the user's head position) or blended between world-locked and head-locked. The device and/or application should support interaction methods that allow the user to make sense of the content.	Add item - adjust size of objects & functionality of fitting obj within a given environment so there's flexibility in interactions. Ask if it works the way it's intended for ease of use	Reword example	The device and/or application should be able to adapt virtual elements in a useful manner. If the virtual content needs to be world-locked, allowing the user to walk around it, the device and/or application should support this. This also applies to content that should be head-locked (follows the user's head position) or blended between world-locked and head-locked. The device and/or application should support interaction methods, such as rescaling the environment, that allow the user to make sense of the content.

Discussion. Experts reported that the experimental heuristic checklist was comprehensive, easy to use, and that it would be helpful in their field. Some additions and edits were suggested to improve the heuristic checklist as described in Table 14. The results from this step were used in step 8 (refinement stage) to determine if checklist items should be kept, changed, added, or removed to revise the experimental heuristic checklist.

Heuristic Evaluation

Evaluators conducted heuristic evaluations on the five applications described in the beginning of the step 7 section in Table 11 (a variety of local applications using the Epson

Moveiro BT-300, Wayfair Spaces using the Magic Leap 1, and Fan Blade Replacement application using the Magic Leap 2, ShapesXR using the Meta Quest Pro, and Google Maps Live View using mobile phones (iPhone 13 Pro or Android Galaxy S8)). Each of these are described in detail in Appendix G, and the technical specifications for each of the devices are described in Table 4.

All heuristic evaluators were Human Factors students from Embry-Riddle Aeronautical University and had experience conducting heuristic evaluations. For each application, three evaluators conducted evaluations with the experimental heuristic checklist, and three additional evaluators conducted evaluations with the control heuristic checklist (de Paiva Guimarães & Martins, 2014). This heuristic checklist was used as a control because it was a validated heuristic checklist for AR and MR applications that covered a large amount of AR features and usability attributes It also did not depend on a specific device or application type to be used. Some items were slightly altered to generalize the terminology across different technologies (e.g., "Is it easy to remember the application's functionalities? (i.e., is it easy to memorize the functionalities of each marker?)" to "Is it easy to remember the application's functionalities? (i.e., is it easy to memorize the functionalities of each button or gesture?)"). Heuristic evaluators were given specific tasks to complete with their assigned application, described in Appendix G, and were provided with a Microsoft Excel document with their randomly assigned heuristic checklist.

The results from the heuristic evaluations conducted with the experimental heuristics were checked against the results from the evaluations conducted with the control heuristics. This addressed research questions 1-3:

1. Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more usability/UX issues** than the use of a control heuristic checklist?
2. Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more issues that qualify as severe/critical** than the use of a control heuristic checklist?
3. Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more domain specific issues** than the use of a control heuristic checklist?

Inter-rater reliability was also conducted to assess how similarly evaluators rated the same application on each heuristic.

Results. More usability/UX violations were found with the experimental heuristic checklist than the control heuristic checklist. There were also more usability/UX violations that qualified as critical found using the experimental heuristic checklist than the control heuristic checklist. The same was found for domain specific issues. This is shown for each application in Table 15.

Table 15

Number of Heuristic Violations for Each Checklist and Device/Application by Type of Violation

	Checklist					
	Control Heuristics (de Paiva Guimarães & Martins, 2014)			**Experimental Heuristics**		
Device & Application	**Usability/UX**	**Domain Specific**	**Critical**	**Usability/UX**	**Domain Specific**	**Critical**
Epson Moverio BT-300 & Variety of Apps Heuristic Evaluations	13	3	2	54	26	16
Magic Leap 1 & Wayfair Spaces	7	3	1	38	17	11
Magic Leap 2 & Fan Blade Replacement	14	6	1	56	21	13
Meta Quest Pro & ShapesXR	11	2	2	65	24	19
Mobile Phone & Google Maps Live View	9	3	1	46	15	12

Note: A violation was defined as an item rating of a "no" or "somewhat" on the heuristic checklist.

Inter-rater reliability was tested using Krippendorff's alpha, where alpha > 0.8 is considered strong, 0.67 to 0.8 is considered low, and < 0.67 is considered really low (Hayes & Krippendorff, 2007). All the alpha scores for both the experimental and control heuristic checklists were really low (<0.67). For every application that was evaluated, alpha levels were higher for the experimental heuristic checklist than the control. This meant that evaluators who were in the experimental heuristic checklist group rated the applications more similarly to each other than those who were in the control heuristic checklist group did. A summary of alpha scores is shown in Table 16.

Table 16

Krippendorff's Alpha for Inter-Rater Reliability of the Heuristic Evaluations

	Checklist	
	Control Heuristics (de Paiva Guimarães & Martins, 2014)	Experimental Heuristics
Device & Application	α	α
Epson Moverio & Variety of Apps Heuristic Evaluations	0.072	0.4457
Magic Leap 1 & Wayfair Spaces	0.1811	0.4774
Magic Leap 2 & Fan Blade Replacement	0.0999	0.3975
Meta Quest Pro & ShapesXR	0.1502	0.1909
Mobile Phone & Google Maps Live View	0.1517	0.2817

Note: > 0.8 is considered strong, 0.67 to 0.8 is considered low, and < 0.67 is considered really low (Hayes & Krippendorff, 2007).

Feedback about the experimental heuristic checklist as a whole and individual items received from evaluators resulted in 12 changes. This included rewording items and examples and adding new items. For example, one heuristic evaluator stated that it was unclear if the item "Does the device adjust to the environment it is used in?" in the Consistency & Standards heuristic meant automatic adjustment or manual adjustment by the user. This item and example were reworded in order to clarify this item. The new item states, "Does the device allow for adjustment based on the environment it is being used in?" and the new example states "E.g., the screen automatically dims in a darker environment, the user can choose to dim screen manually, the volume adjusts when the user is in a loud environment, etc." Two items were added, two examples were added, two items were reworded, and 6 examples were reworded. These changes are described in detail in Table 17.

Table 17

Changes Made from the Feedback Given by the Heuristic Evaluators

Heuristic	Original Checklist Item/Example	Feedback	Change(s) made	Updated Item/Example
Consistency & Standards	Can the user pause the application at any point? The user should be able to easily stop what they are doing and return to their same spot in the application as they please. If there is an interruption in the real-world environment that the user has to attend to, they should **NOT** have to turn off and/or restart the application and lose their progress.	This question is confusing. I can take off the headset and return back to the app later, but there is not function called "pause"	Reword Example	Terminology was added to the example: "E.g., include a "pause" button, automatically pause when user presses "home" button or when they take off the device"

Heuristic	Original Checklist Item/Example	Feedback	Change(s) made	Updated Item/Example
Consistency & Standards	Item: Does the device adjust to the environment it is used in? Example: e.g., does the screen dim in a darker environment?	I am unsure if this means auto adjustment, this doesn't do it automatically but does manually	Reword item and example	Item: Does the device allow for adjustment based on the environment it is being used in? Example: E.g., the screen automatically dims in a darker environment, the user can choose to dim screen manually, the volume adjusts when the user is in a loud environment, etc.
Collaboration	Item: When collaborating with others, is a private space for the user provided? Example: It may be necessary for a user to understand what "I" see vs. "others" can see. Users may not want to share all of their virtual elements with others or may want to view it before sharing with others. A private space can enable this interaction without sharing all content to other collaborators.	Reword item to make it clearer	Reword item and example	Item: When collaborating with others, is it clear what content is and is not private? Terminology was added to the example: "It should be clear to the user what information is and is not private and how to alter those settings to make it less/more secure as necessary. I.e., Enabling a private space where the user can see and change content without sharing all content to other collaborators."
Feedback	Does the device and/or application provide feedback on its status? e.g., loading screen, representation when scanning the environment, etc.	The device's connection seemed to cut out a few times and you only know because the other collaborator disappears, which made me think that I did something wrong. A connection issue message would have helped me understand the problem faster.	Reword example	does the device/app provide feedback on its status? e.g., loading screen, visual representation when scanning the environment, notification presented when there are connection issues, etc.

Heuristic	Original Checklist Item/Example	Feedback	Change(s) made	Updated Item/Example
Integration of Physical & Virtual Worlds	N/A	The app did not show most efficient route	Add item	Do the virtual elements help the user accomplish the required tasks in a meaningful way?
Integration of Physical & Virtual Worlds	N/A	Is the app useful vs. novel? Does it cover a need? Like try in your space vs. just scrolling through pictures/reviews in amazon.	Add item (same as item above)	Do the virtual elements help the user accomplish the required tasks in a meaningful way?
Integration of Physical & Virtual Worlds	N/A	Some of the instructions for the task steps were not provided, like replacing the screws once the blade was changed.	Add item (same as item above)	Do the virtual elements help the user accomplish the required tasks in a meaningful way?
Privacy	N/A	I didn't like that you had to point your camera at the environment all the time, potentially creeping other people out. Maybe could ask if the app is intrusive, or something like that?	Add item	Does the device and/or application avoid those around the user to the utmost that the required tasks allow?

Heuristic	Original Checklist Item/Example	Feedback	Change(s) made	Updated Item/Example
User Interaction	Example: The device and/or application should be able to adapt virtual elements in a useful manner. If the virtual content needs to be world-locked, allowing the user to walk around it, the device and/or application should support this. This also applies to content that should be head-locked (follows the user's head position) or blended between world-locked and head-locked. The device and/or application should support interaction methods that allow the user to make sense of the content.	There is no resizing option for the virtual room so that I could see the object within it better	Reword Example	Terminology was added to the example: "The device and/or application should support interaction methods, such as rescaling the environment, that allow the user to make sense of the content."
User Interaction	Examples: "Allowing the user to interact how it works best for them makes a more seamless interaction experience, this can be done by implementing multiple forms of input (e.g., mouse & keyboard input, taps, gesture controls, voice commands, etc.)…" If there are multiple ways to interact with an object (e.g., "select" an object through a voice command, one-handed pinch motion, two-handed gesture, gaze, or menu option), ensure that these interactions work well in all instances (e.g., if user is walking, interacting with another object, holding a mobile device with one hand, etc.).	There are no heuristics for head-motion input	Reword Examples	Terminology was added to the examples: "(e.g., mouse & keyboard input, taps, gesture controls, voice commands, head-motion input, etc.)" "If eye or head gaze is used, it is recommended to use delay timers and dwell for selection to ensure that a user's input is intentional rather than accidental."

Discussion. This portion of the validation process compared the experimental heuristic checklist to another heuristic checklist that is being used in this domain. The goal was to investigate whether the experimental heuristic checklist resulted in more usability/UX issues, issues that qualify as severe/critical, domain specific issues, and inter-rater reliability when compared to the control heuristic checklist. Overall, evaluators using the experimental heuristic checklist found more violations in each of these categories for each application evaluated. The experimental checklist included the categories of help & documentation, comfort, and user interaction that all qualified as critical. These categories were not as extensively covered in the control checklist

Inter-rater reliability was tested using Krippendorff's alpha, where alpha > 0.8 is considered strong, 0.67 to 0.8 is considered low, and < 0.67 is considered really low (Hayes & Krippendorff, 2007). All the alpha scores for both the experimental and control heuristic checklists were really low (<0.67). This is consistent with previous studies on heuristic evaluations (Smith, 2021; Leverenz, 2019; and White et al., 2011). After tasks were completed, heuristic evaluators were given the chance to explore the application as needed as they completed the evaluation. Some evaluators may have explored the application more than others, and as a result spent more time on their evaluations. This could be why inter-rater reliability was very low for both heuristic checklists. Additionally, heuristic evaluations are subjective methods for evaluating a product. It is possible that heuristic evaluators may have different opinions of the product, and as a result, evaluate it differently. For example, the following question in the experimental checklist asks about the evaluator's personal experience, and as a result different evaluators may answer this question differently, "Can the user experience the device and/or application without pain, discomfort, nausea, disorientation, etc. DURING use?”. For every application that was evaluated, alpha levels were higher for the experimental heuristic checklist than the control. The higher the alpha level, the larger the inter-rater reliability was between evaluators. Since the experimental heuristic checklist scored higher, this means that evaluators rated heuristic items more similarly to other evaluators in this group than the control evaluators did with others in their group. This showed that the evaluations conducted with the experimental heuristic checklist were more reliable than those conducted with the control heuristic checklist, however, still provide variability between raters. As a result, to identify the most amount of usability issues as possible, it is recommended that multiple raters conduct heuristic evaluations and discuss their results together to create recommendations. This recommendation is consistent with other heuristic checklists (Nielsen, 1994b).

Heuristic evaluators also provided feedback about the heuristic checklist. This led to meaningful changes for the final version of the experimental heuristic checklist. Items were revised, new examples were made, and some items were created as a result from the evaluators’ feedback.

User Testing

Five user tests were conducted, each with a different device and application. The results from the user tests were compared to the results found during the experimental heuristic checklist evaluations. This was done to answer the fourth research question:

4. Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **similar usability/UX issues as a usability study** of the same application?

Information about each of the applications and devices are shown in Tables 4, 11, and Appendix G. Each of the five user tests used the following common materials:

Task Success and Difficulty. Task success rates were documented by noting if a participant successfully completed a task or not. Task difficulty was self-reported by participants after they completed each assigned task on a scale of 1-5 (1 = very easy, 5 = very difficult). Participants were also asked how difficult they expected it to be to use the device before completing tasks, and how difficult it was to use the device after completing all tasks using this same scale.

System Usability Scale (SUS). The SUS is a 10-item questionnaire that is used to assess the usability of the application (Brooke, 1996). Participants were asked to complete this questionnaire after they completed all tasks with the device.

NASA-Task Load Index (NASA-TLX) Raw. The NASA-TLX (Hart & Staveland, 1988) is a 6-item questionnaire that is used to assess the perceived workload of a task or system. Participants are asked to rate their workload from a scale of 1 to 21 on six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration. This study used the NASA-TLX raw values, rather than weighted values, in order to assess how a participant felt on each of these subscales. Participants were asked to complete this questionnaire after they completed all tasks with the device.

Simulator Sickness Questionnaire (SSQ). The SSQ (Kennedy et al., 1993) is a 4-point Likert-scale (none (0), slight (1), moderate (2), and severe (3)) 16-item questionnaire that assesses the amount of perceived sickness a participant is feeling after using a system. It includes an overall score and scores from three sub-factors: nausea, oculomotor discomfort, and disorientation. A categorization of symptoms for a simulator can range from a score of 0 (no symptoms) to a score larger than 33.3 (extreme symptoms) (Stanney et al., 2021). Participants were asked to complete this questionnaire after they completed all tasks with the device.

Eyestrain Questionnaire. Six questions 5-point Likert-scale questions were asked to assess a participant's level of eye strain. These were coded so the most negative rating was 1 and most positive rating was 5. These included: ease of reading text (1 = very difficult, 5 = very easy), text clarity (1 = very dissatisfied, 5 = very satisfied), ability to concentrate (1 = very low, 5 = very high), physical fatigue (1 = very high, 5 = very low), mental fatigue (1 = very high, 5 = very low), and level of eyestrain (1 = very high, 5 = very low). Participants were asked to complete this questionnaire after they completed all tasks with the device.

User Test 1: Epson Moverio & a Variety of Applications.

Materials. The measures described above (task success and difficulty, SUS, NASA-TLX, SSQ, and eyestrain questionnaire) were used for this study. Participants were also asked to complete the same tasks that heuristic evaluators completed with the Epson Moverio BT-300. These are described in detail in Appendix G.

Participants. Participants included 8 college students from Embry-Riddle Aeronautical University and were recruited from either a university online research participation system or from word of mouth. All participants had either normal vision or corrected to normal vision by wearing contact lenses. The participants were comprised of one male and seven females with ages ranging from 20 to 29 (M = 23.75, SD = 3.37). Two participants were left-handed while six were right-handed. Seven participants stated that they had used XR devices before (five of which have used AR or MR devices before). These previous experiences were often short (one participant used XR devices for a total of less than one hour; three used XR devices for a total of 1-4 hours; one used XR devices for a total of 5-9 hours; one participant used XR devices for a total of 10-19 hours, and one participant who owned a Meta Quest 2 (VR device) at home used XR devices for a total of 40+ hours. Six participants reported playing video games regularly (4 of which self-identified as a "casual" gamer, and 2 self-identified as a "hardcore/expert" gamer). Five participants reported playing video games between 1-4 hours per week and one participant reported playing video games between 10-19 hours per week. All participants were asked to give their informed consent before beginning the study.

Procedure. The procedure for this user test is described in detail in Appendix G. In summary, participants were asked to play games, take and review photos, and watch a video on the Epson Moverio BT-300. Participants gave their opinions on their satisfaction with the application and device, how difficult it was to complete tasks, and completed the questionnaires described earlier in this section (SUS, NASA-TLX Raw, SSQ, and Eyestrain Questionnaire).

Results. Detailed results regarding the task success and difficulty, SUS, NASA-TLX, SSQ, and eyestrain questionnaire questionnaires can be found in Appendix H. Overall, it was found that the Epson Moverio smart glasses device was lightweight and overall simple to use. However, it caused a lot of discomfort for participants, mostly because the fit was not adjustable for different users. There was no tutorial for the device, but it was simple to use and easy to learn. Head controls were uncomfortable to complete, but more accurate than the remote controls. There was a lack of integration with the real world, which participants found made the device less valuable and more difficult to focus on when there was a busy background (background filled with computers, posters, etc.). Only half of participants (4/8) would use it again in niche use cases (training or teaching tool). A summary of usability issues found as a result of the user test is in Table 18.

Table 18

Epson Moverio BT-300: Usability Issues Identified During the User Test

Usability Issue	Description
Placement of buttons make it difficult to use the controller	The remote included quick action buttons (return to home, back, see all apps). Participants often accidentally slid their thumb up too far and hit a quick action button exiting the app.
Typing is tedious	Participants thought that typing was tedious and took up too much time. They had to scroll and manually select each letter on a virtual keyboard to type.
Head movement controls are disorientating	Participants believed they had to make grand head movements to control some of the apps. These significant movements, and the fact that the AR screen moved as they moved their head, caused some dizziness and discomfort.
Not enough user control in Asteroid Fighter	During the Asteroid Fighter game, participants controlled a spaceship and destroyed asteroids. They could control the pitch, roll, and yaw but not the speed of the spaceship. They stated this was frustrating to have a lack of control that they expected to be given.

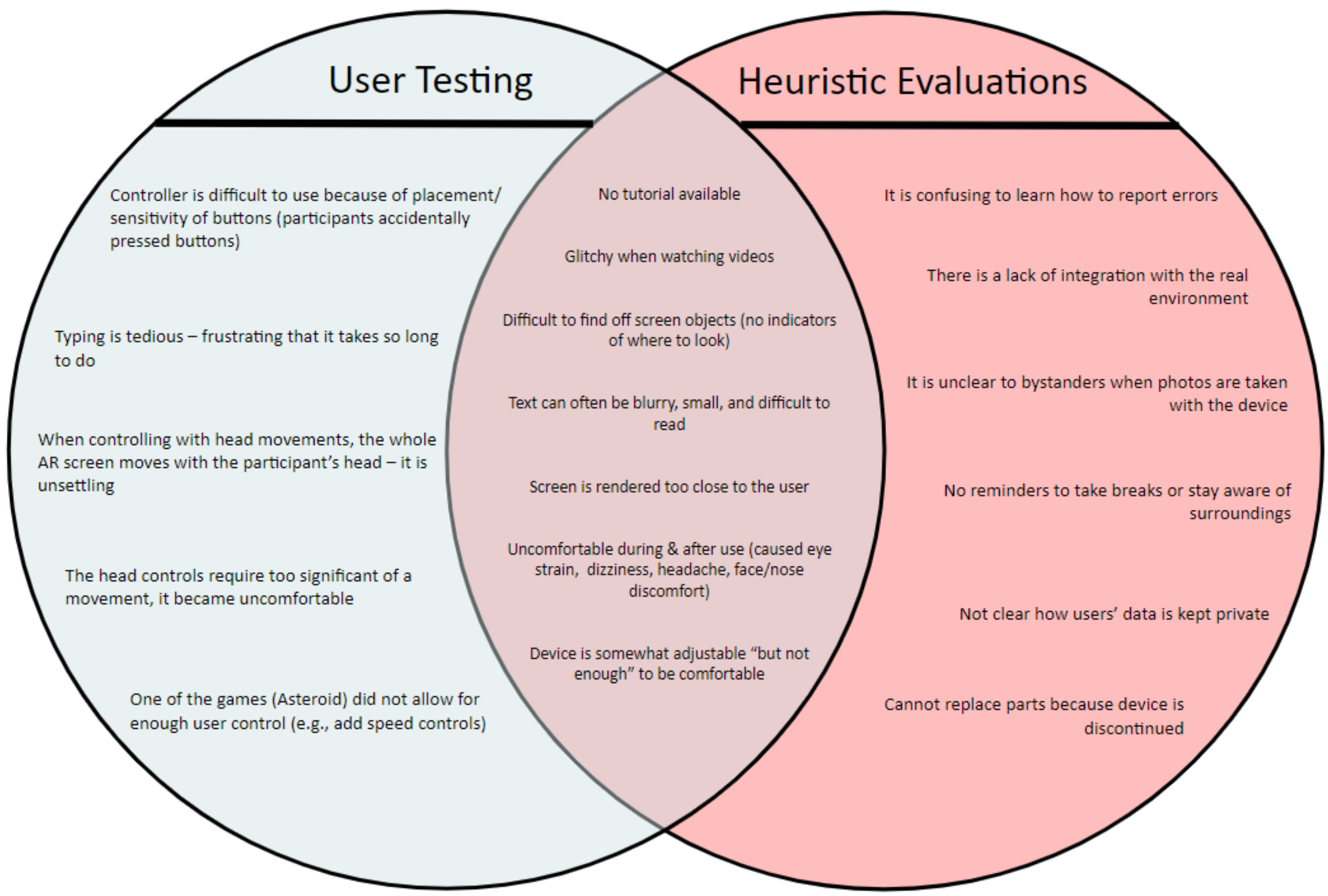
Usability Issue	Description
No tutorial	A quick start guide was provided to users, but no tutorials were integrated into any of the apps. This confused users about how to use game controls and mechanics.
Jitter/Glitches	Button presses did not respond to user input reliably. Sometimes a button press would result in a change on the screen, sometimes it did not.
No indication to where off-screen objects are located	During the start of the Age of Diamonds game, participants had to find the "Level 1" button which was often located off-screen. It took a large amount of time to find this virtual object because there were no cues letting them know where the icon was located.
Text is difficult to read	Text was blurry, small, and difficult to read.
AR content is rendered too close to user	The AR screen was automatically rendered in a specific spot for users. Many stated that it was too close and caused their eyes to become strained. This could not be adjusted.
Hardware is not sufficiently adjustable to fit a variety of users, making it uncomfortable to wear	Participants felt uncomfortable wearing the device due to the lack of customized fit (no eye calibration, only 2 different nose pieces, no ability to make the device tighter).

Discussion. These smart glasses are a very lightweight and simple way to introduce AR to new users. However, participants reported a lot of discomfort (mostly because the fit was not adjustable for different users). There was no tutorial for the device, but it was simple to use and easy to learn. Head controls were uncomfortable to complete, but more accurate than the remote controls. There was a lack of integration with the real world, which users found made the device less valuable and more difficult to focus on when there was a busy background (background filled with computers, posters, etc.). Only half of participants (4 out of 8) would use it again in niche use cases (e.g., training or a teaching tool). This user test identified key usability issues with the application.

Results from this user test were compared to results of the heuristic evaluations of the same application. Many of the usability issues were found using both methods, especially those that involved comfort of the device, quality of the screen and audio, instructions. Issues that were more subjective to each user's personal experience were found mostly during user testing, such as whether the device was "cool", interactions that were "unsettling" or "tedious", and the wish that they had more interactions in a game that was played. Issues that were about long-term use (e.g., replacement of parts, how to report errors, reminders to take breaks, etc.) or how the device impacts privacy were mostly found in heuristic evaluations. A summary of these findings is shown in Figure 12.

Figure 122

Epson Moverio BT-300: Comparing Usability Issues Found in User Testing and Heuristic Evaluations



User Test 2: Magic Leap 1 & Wayfair Spaces.

Materials. The measures described above (task success and difficulty, SUS, NASA-TLX, SSQ, and eyestrain questionnaire) were used for this study. Participants were also asked to complete the same tasks that heuristic evaluators completed with Wayfair Spaces application on the Magic Leap 1. These are described in detail in Appendix G.

Participants. Participants included 9 college students from Embry-Riddle Aeronautical University and were recruited from either a university online research participation system or from word of mouth. All participants had either normal vision or corrected to normal vision by wearing contact lenses. The participants were comprised of four males and five females with ages ranging from 18 to 25 (M = 21.33, SD = 2.06). One participant was left-handed while eight were right-handed. Seven participants stated that they had used XR devices before (three of which have used AR or MR devices before). These previous experiences ranged vastly between participants (three participants used XR devices for a total of 1-4 hours; one participant used XR devices for 20-29 hours, and three used XR devices for 40+ hours. Three out of the four participants who used XR devices for 20-29 hours and 40+ hours reported owning personal VR headsets at home. Six participants reported playing video games regularly (1 of which self-identified as a "newbie/novice" gamer, 3 self-identified as a "casual" gamer, and 2 self-identified as a "mid-core" gamer). One participant reported playing video games less than 1 hour per week, two reported playing between 1-4 hours per week, two reported playing video games between 10-19 hours per week, and one reported playing videogames between 20-29 hours per week. All participants were asked to give their informed consent before beginning the study.

Procedure. The procedure for this user test is described in detail in Appendix G. In summary, participants were asked to complete a tutorial and add virtual furniture into their real space using the Wayfair Spaces Application and the Magic Leap 1. Participants gave their opinions on their satisfaction with the application and device, how difficult it was to complete tasks, and completed the questionnaires described earlier in this section (SUS, NASA-TLX Raw, SSQ, and Eyestrain Questionnaire).

Results. Detailed results regarding the task success and difficulty, SUS, NASA-TLX, SSQ, and eyestrain questionnaire questionnaires can be found in Appendix H. Overall, users had a very positive experience with this application (6 of 9 participants said they would use this application again). Most of the issues that participants had were a result of how the tutorial portrayed information. For example, participants did not remember how to find objects from the items tab or another scene because the tutorial only told them how to do this through text on the screen instead of asking users to practice the task like the other parts of the tutorial. Overall, this application is very usable, but participants wanted more control over the things they do, some tasks need to be more straightforward and efficient, and glitches need to be fixed. A summary of usability issues found as a result of the user test is in Table 19.

Table 19

Magic Leap 1: Usability Issues Identified During the Wayfair Spaces User Test

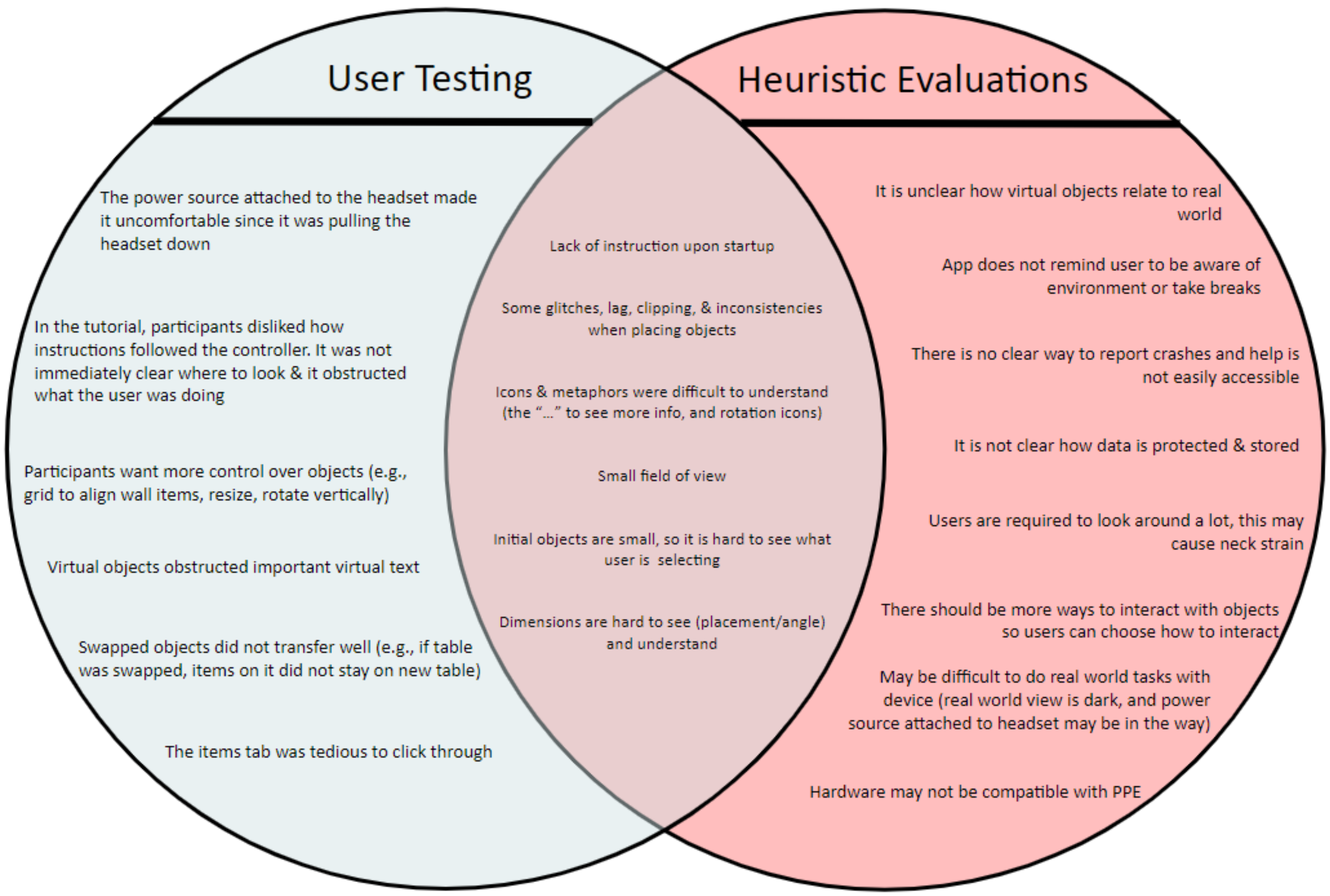
Usability Issue	Description
Lack of instruction upon start up	When participants started the app, they saw three 3D renderings of rooms and no additional information. They were confused how to start the app and move on from this page due to a lack of instruction.
MR content tethered to the controller obstructed view and was often off-screen	Tutorial instructions were tethered to the controller rather than on-screen UI. As a result, when participants tried to complete a task, the instructions would be in their view and sometimes obstructing their task. Sometimes participants' controller would be off-screen, instructions would change, and they were unaware.
MR content obstructed other content	Sometimes MR objects would obstruct text that the participants needed to read.
MR content did not act as expected	When changing a MR object to something else (e.g., changing a tall bookcase into a short bookcase) items that were on top of that original MR object would not adjust (e.g., a plant that was on top of the tall bookcase would float in air when the MR object was swapped to the short bookcase).
Menu was tedious to click through	Participants had to click through many menu options to get to what they needed.
Jitter/Glitches	Some MR objects, especially wall hangings, glitched/clipped into the real wall.
Icons and metaphors used are not common	Some participants had difficulties understanding what icons mean (e.g., the "..." icon) because their form did not communicate their function.
Some MR objects are too small or difficult to see	Participants had difficulties selecting/moving small MR objects with their controller. Some text was also too small to read.

Discussion. Participants overall had a pleasant experience with this application. There are some glitches, but most of the problems had to do with how the tutorial portrayed information (users did not remember how to find objects from the items tab/another scene because they did not practice it in the tutorial). Overall, this application is designed well but users want more control over the things they do, some tasks need to be more straightforward and efficient, and glitches need to be fixed. Six out of nine participants stated that they would use this application again in the future. This user test identified key usability issues with the application.

Results from this user test were compared to results of the heuristic evaluations of the same application. Many of the usability issues were found using both methods, especially those that involved quality of the screen and audio, instructions, the organization of the user interface, and ability to see virtual holograms. Issues that were more subjective to each user's personal experience were found mostly during user testing, such as whether the device or application was "cool". It was also noted that some interactions were "tedious", and participants wished that they had more control over objects within the app. Issues that were about long-term use (e.g., how to report errors, reminders to take breaks, etc.), how the device impacts privacy, and completing real-world tasks with the device on were mostly found in heuristic evaluations. A summary of these findings is shown in Figure 13.

Figure 13

Magic Leap 1: Comparing Usability Issues Found in User Testing and Heuristic Evaluations



User Test 3: Magic Leap 2 & Fan Blade Replacement.

Materials. The measures described above (task success and difficulty, SUS, NASA-TLX, SSQ, and eyestrain questionnaire) were used for this study. Participants were also asked to complete the same tasks that heuristic evaluators completed with the Magic Leap 2 and the Fan Blade Replacement application. These are described in detail in Appendix G.

Participants. Participants included seven college students from Embry-Riddle Aeronautical University and were recruited from either a university online research participation system or from word of mouth. All participants had either normal vision or corrected to normal vision by wearing contact lenses. The participants were comprised of four males and three females with ages ranging from 20 to 28 (M = 23.29, SD = 3.15). One participant was left-handed while six were right-handed. All seven participants stated that they had used XR devices before (all participants stated they used AR or MR devices before). These previous experiences ranged vastly between participants (three participants used XR devices for a total of 1-4 hours; two participants used XR devices for 10-19 hours, and two used XR devices for 40+ hours. Both participants who stated they have used XR devices 40+ hours reported owning personal VR headsets at home. All seven participants reported playing video games regularly (2 of which self-identified as a "casual" gamer, 3 self-identified as a "mid-core" gamer, and 2 self-identified as a "hardcore/expert" gamer). Two participants reported playing between 1-4 hours per week, two reported playing video games between 5-9 hours per week, two reported playing between 10-19 hours per week, and one reported playing between 20-29 hours per week. All participants were asked to give their informed consent before beginning the study.

Procedure. The procedure for this user test is described in detail in Appendix G. In summary, participants were asked to follow instructions given to them by holograms on the Magic Leap 2 to help guide them through the task of changing a fan blade on an aircraft engine. Participants gave their opinions on their satisfaction with the application and device, how difficult it was to complete tasks, and completed the questionnaires described earlier in this section (SUS, NASA-TLX Raw, SSQ, and Eyestrain Questionnaire).

Results. Detailed results regarding the task success and difficulty, SUS, NASA-TLX, SSQ, and eyestrain questionnaire questionnaires can be found in Appendix H. At first glance, participants said the application was eye-catching and exciting to use. Participants were excited that they were able to get hands-on experience learning a task. However, after using the application, participants encountered tasks that made it difficult to change the fan blade and they began to lose trust in the application's instructions. This occurred during the following tasks: replace fan blades, replace rear segment, and replace front segment. These tasks should be edited to help users complete the procedure. Most participants (5 out of 7) said they would use this application over written or video instructions. A summary of usability issues found as a result of the user test is in Table 20.

Table 20

Magic Leap 2: Usability Issues Identified During the Fan Blade Replacement User Test

Usability Issue	Description
Participants were unsure how to interact with the app at first (no tutorial provided)	No tutorial was provided upon first use of the app. As a result, participants did not know how to interact with the app (use voice commands to move forward).
Lack of consistency within app	Similar tasks were completed at the beginning and the end, but were asked to be completed differently. For example, tasks 1 & 2 are two separate tasks that were later combined into one (task 11) Task 1 – loosen bolts on the front cover & Task 2 – remove the front cover; Task 11 – place and secure the front cover.
App did not act as expected	Participants wanted to repeat MR animations to watch what they needed to do to the engine. They though "repeat task" would repeat animations but it did not.
Some MR visuals and terminology were unclear	It was sometimes unclear what real object the MR content was referencing. E.g., an "indentation" described in MR was not an indentation nor in the same location in the real environment.

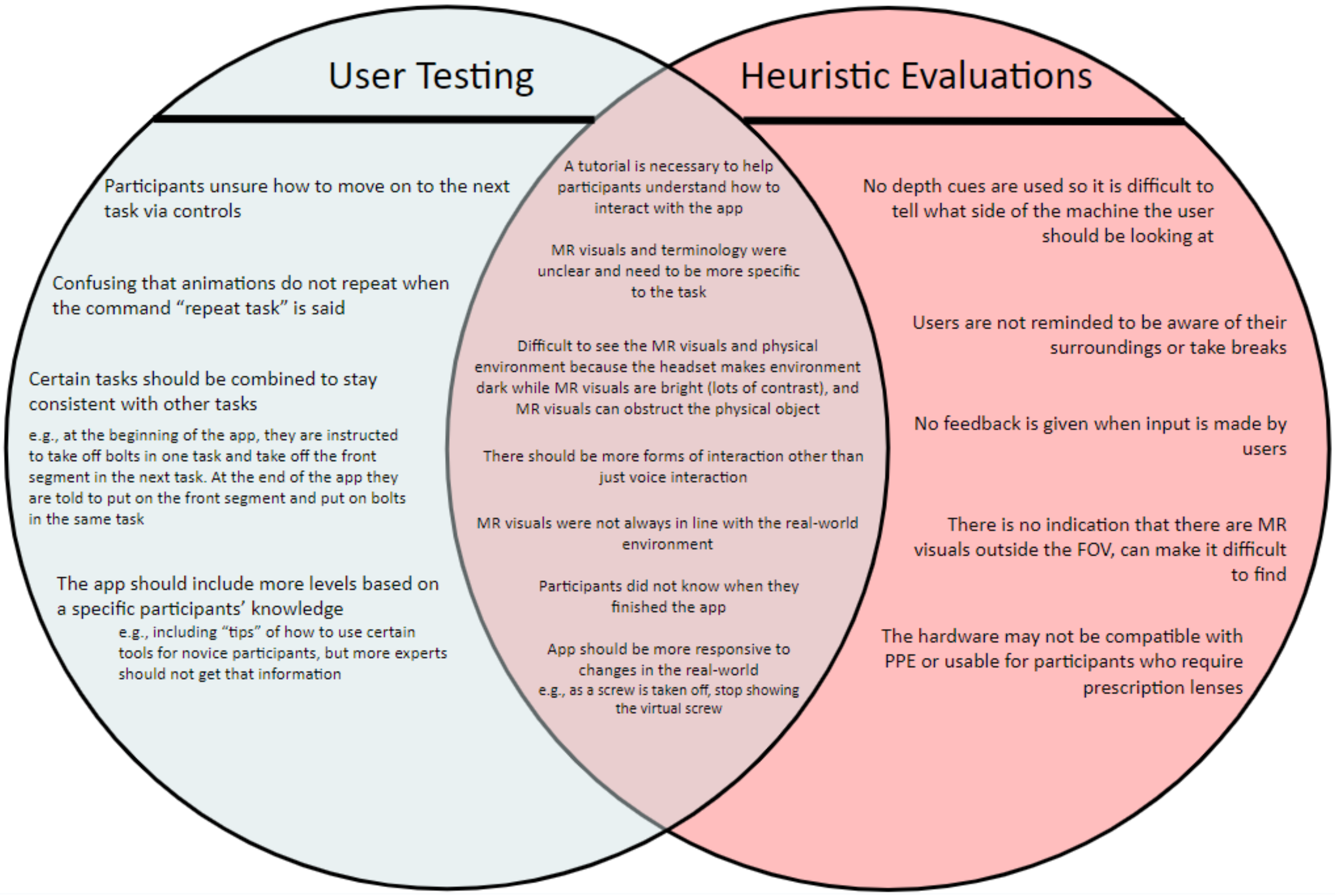
Usability Issue	Description
MR content did not seamlessly integrate with the real world	MR visuals were not always aligned with the real-world reference. Additionally, during some tasks, participants had a hard time seeing the real environment while wearing the MR device. This is because the MR objects were very bright while the real environment was very dark. A light was sometimes necessary to see and interact with real objects.
No multiple forms of interaction	Participants could interact with the app using voice commands, but this did not always work (especially when the environment was loud). Gestures only moved the instructions and could not be used to continue to the next task. No other forms of interaction were available to participants.
Participants were unsure when they finished all of the tasks	After completing all tasks, there was no indication that the participant was "done". Many were confused if they completed the tasks correctly or not for this reason.

Discussion. At first glance, the application was eye-catching and exciting to use. Participants were excited that they were able to get hands-on experience. However, after using the application, users encountered tasks that made it difficult to change the fan blade and they began to lose trust in the app's instructions (Replacing fan blades, replace rear segment, and replace front segment). These tasks should be edited to help users complete the procedure. Most participants (5 out of 7) said they would use this application over written or video instructions. This user test identified key usability issues with the application.

Results from this user test were compared to results of the heuristic evaluations of the same application. Many of the usability issues were found using both methods, especially those that involved instructions, how intuitive the application was, and how the AR content was integrated into the real environment. Issues that were more subjective to each user's personal experience were found mostly during user testing, such as whether the application was "cool" and a lack of knowledge the participants had with maintenance tasks (none of the participants were aircraft maintenance students). It was also noted that it was unclear how to initially use the application (possibly due to the lack of experience participants had with AR before this study), and that consistency between tasks should be improved. Issues that were about long-term use (e.g., how to report errors, reminders to take breaks, etc.), concerns about hardware compatibility with personal protective equipment (PPE), and amount of feedback given to the user were mostly found in heuristic evaluations. A summary of these findings is shown in Figure 14.

Figure 144

Magic Leap 2: Comparing Usability Issues Found in User Testing and Heuristic Evaluations



User Test 4: Meta Quest Pro & ShapesXR.

Materials. The measures described above (task success and difficulty, SUS, NASA-TLX, SSQ, and eyestrain questionnaire) were used for this study. Participants were also asked to complete the same tasks that heuristic evaluators completed with the Meta Quest Pro and ShapesXR application. These are described in detail in Appendix G.

Participants. Participants included 8 college students from Embry-Riddle Aeronautical University and were recruited from either a university online research participation system or from word of mouth. All participants had either normal vision or corrected to normal vision by wearing contact lenses or glasses. The participants were comprised of one male and seven females with ages ranging from 23 to 38 (M = 27.63, SD = 5.71). One participant was left-handed while seven were right-handed. All eight participants had previous experience with 2D prototyping applications (e.g., Figma, Axure, Balsamiq, etc.). All eight participants stated that they had used XR devices before (five of which have used AR or MR devices before). The amount of time spent during these previous experiences ranged between participants (one participant used XR devices less than one hour, three participants used XR devices for a total of 1-4 hours, one used XR devices for 5-9 hours, one used XR devices for 10-19 hours, one used XR devices for 20-29 hours, and one used XR devices for 40+ hours). One participant reported owning a personal VR headset at home, but had limited experience using XR devices (only 10-19 hours of experience). Seven participants reported playing video games regularly (2 of which self-identified as a "newbie/novice" gamer, 3 self-identified as a "casual" gamer, 1 self-identified as a "mid-core" gamer, and 1 self-identified as a "hardcore/expert" gamer). Two participants reported playing video games less than 1 hour per week, two reported playing between 1-4 hours per week, one reported playing between 5-9 hours per week, one reported playing 10-19 hours per week, and one reported playing videogames 40+ hours per week. All participants were asked to give their informed consent before beginning the study.

Procedure. This is described in detail in Appendix G. In summary, participants were asked to complete tutorials and build a prototype with another user (a confederate researcher) with the ShapesXR application using the Meta Quest Pro. Participants gave their opinions on their satisfaction with the application and device, how difficult it was to complete tasks, and completed the questionnaires described earlier in this section (SUS, NASA-TLX Raw, SSQ, and Eyestrain Questionnaire).

Results. Detailed results regarding the task success and difficulty, SUS, NASA-TLX, SSQ, and eyestrain questionnaire questionnaires can be found in Appendix H. Overall, participants felt very immersed and thought the application was fun. They also saw it as a useful application (6 of 8 participants said they would use the application again). But, the number of controls was overwhelming. For example, participants forgot how to complete some "simple" tasks, as they described them, because they could not remember which button to push to access the control. Collaborating with a partner went well, mostly due to the presence of an avatar that provided non-verbal cues, and participants were able to complete tasks with their partner. Participants reported that the tutorials need an overhaul to help users ease into the application. The tutorial videos that participants watched included a large amount of information in a short video, no audio, a lack of subtitles, no context, and did not provide an opportunity for users to practice what they learned. Because of this, users felt overwhelmed after tutorials and were inaccurate during the tasks. The tutorial should engage the user more to help them retain the information. A summary of usability issues found as a result of the user test is in Table 21.

Table 21

Meta Quest Pro: Usability Issues Identified During the ShapesXR User Test

Usability Issue	Description
Visibility of real-world while completing tasks	Participants were required to press buttons on the controller to complete tasks, but they were unable to see their real hands over the controller (instead, a button overlay was placed over the real controller, blocking their fingers). As a result, they were unsure which buttons they were pressing.
Teleportation was "jarring"	Participants had a few different options for traveling around the virtual space. One of which was teleporting, but most participants avoided it because it was described as "jarring" since it immediately teleporting the user (rather than transitioning them to the location) and more tedious to aim where to go rather than just walking there.
Uncertainty about user privacy in the collaborative space	Participants were unsure which changes in the environment their collaborator could and could not see. For example, when resizing the virtual space, they were unsure if it resized for their collaborator as well.
Tutorials were overwhelming	Tutorials covered a lot of information in a short period of time, which overwhelmed participants. It was also difficult to remember what was covered in the tutorial because the tutorials showed users what to do (in a muted video) rather than give them context or an opportunity to try what they learned.
Steep learning curve	App controls had a steep learning curve, and it was difficult to remember the controls. This could be due to the amount of information presented in tutorials.
Interacting with virtual objects that were far away was not efficient	Participants could move towards far away objects to interact with them, but they said this took too many steps and would rather interact with it via a raycast or be able to throw virtual objects to their collaborator.

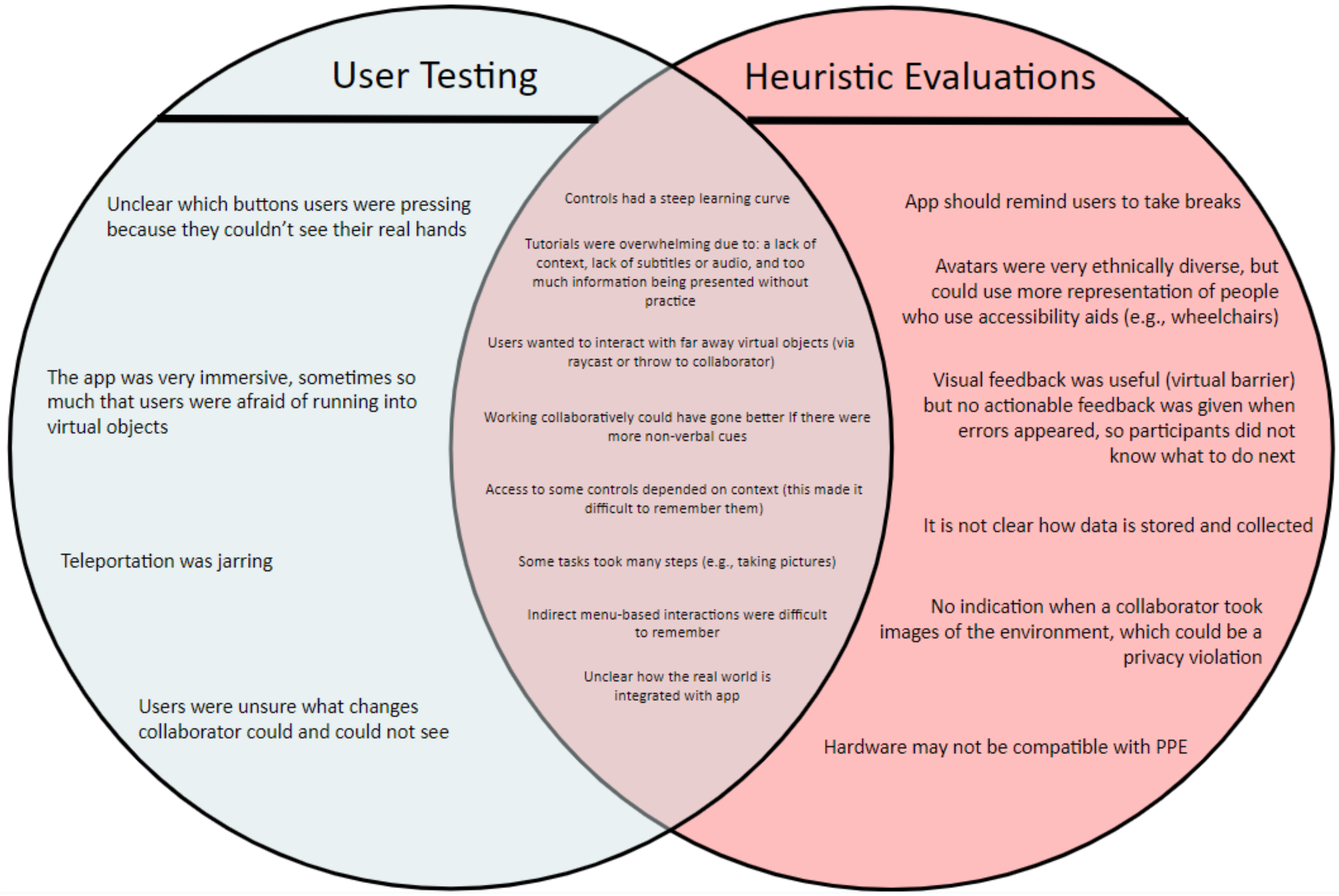
Usability Issue	Description
Controls depended on context	Depending on what the participant was doing in the app, they had access to different controls using the same buttons (e.g., they could “snap” an object only if they were holding it and pressed a button. If they selected the object, but was not holding it, the “snap” function would not be an option). This made it difficult to remember how to complete the action since it was hidden until the participant was in the correct context.
Tasks took many indirect steps	Some of the most difficult tasks (e.g., taking photos and changing the color of objects) required users to remember many steps and menu clicks to complete the task. This was difficult to remember so participants had a difficult time completing these tasks.
Unclear how the real world is integrated with app	Participants said that they were confused how the real world was integrated into the app. They suggested virtual objects to automatically snap to real objects (e.g., sit on top of a desk).

Discussion. Participants felt very immersed and thought the application was fun, but the number of controls was overwhelming. So much so that participants forgot how to complete some tasks. Collaborating with another user went well and participants were able to complete the main task of creating a prototype of an office space. They felt that the application was useful as 6 of 8 participants said they would use it again. This user test identified key usability issues with the application. For example, tutorials need an overhaul to help users ease into the application. Currently, the tutorials only show the user how to use the controls with no context, audio, or practice. The tutorials should be more interactive and ask the user to practice what they learn as they are being instructed.

Results from this user test were compared to results of the heuristic evaluations of the same application. Many of the usability issues were found using both methods, especially those that involved usefulness of the application, instructions, hardware fit, intuitiveness of controls, and cognitive overload. Issues that were more subjective to each user's personal experience were found mostly during user testing, such as whether interactions were "fun" to complete (e.g., "throwing" objects to delete them) or whether the application was immersive. It was also noted that the virtual objects occluded participants real hands, so they had difficulty learning the controls, teleportation was a jarring experience so many participants did not use it, and participants were unsure if changes they made to the MR content could be seen by their partner. Issues that were about long-term use (e.g., how to report errors, reminders to take breaks, etc.), representation of diverse users through avatar creation, how the application impacts privacy, and concerns about hardware compatibility with personal protective equipment (PPE) were mostly found in heuristic evaluations. A summary of these findings is shown in Figure 15.

Figure 155

Meta Quest Pro: Comparing Usability Issues Found in User Testing and Heuristic Evaluations



User Test 5: Mobile Phone & Google Maps Live View.

Materials. The measures described above (task success and difficulty, SUS, NASA-TLX, SSQ, and eyestrain questionnaire) were used for this study. Participants were also asked to complete the same tasks that heuristic evaluators completed with a mobile phone and the Google Maps Live View application. These are described in detail in Appendix G.

Participants. Participants included 9 college students and recently graduated students from Embry-Riddle Aeronautical University and were recruited from either a university online research participation system or from word of mouth. All participants had either normal vision or corrected to normal vision by wearing contact lenses or glasses. The participants were comprised of three males and six females with ages ranging from 22 to 38 (M = 27.56, SD = 4.75). Three participants were left-handed while six were right-handed. All nine participants stated that they had used XR devices before (eight of which have used AR or MR devices before). The amount of time spent during these previous experiences ranged between participants (one participant used XR devices less than one hour, two participants used XR devices for a total of 1-4 hours, two used XR devices for 5-9 hours, one used XR devices for 10-19 hours, one used XR devices for 20-29 hours, and two used XR devices for 40+ hours). Two participants, one who reported having 20-29 hours of experience and one who reported having 40+ hours of experience with XR owned a personal VR headset at home. Eight participants reported using XR apps on mobile devices before (mobile phone or a tablet). Most of these experiences were with retail apps (n=6), games (n=5), social media (n=5), training apps (n=4), educational apps (n=3), and navigation (n=1). Two participants had less than 1 hour of experience with these apps, one had 10-19 hours of experience, four had 20-29 hours of experience, and one had over 40 hours of experience with mobile XR apps. Eight participants reported playing video games regularly (4 self-identified as a “casual” gamer, 3 self-identified as a “mid-core” gamer, and 1 self-identified as a “hardcore/expert” gamer). Five participants reported playing video games between 1-4 hours per week, two reported playing between 5-9 hours per week, and one reported playing 10-19 hours per week. All participants were asked to give their informed consent before beginning the study.

Procedure. The procedure for this user test is described in detail in Appendix G. In summary, participants were asked to walk 0.3 miles on a college campus while following AR directions using the application. Participants gave their opinions on their satisfaction with the application and device, how difficult it was to complete tasks, and completed the questionnaires described earlier in this section (SUS, NASA-TLX Raw, SSQ, and Eyestrain Questionnaire).

Results. Detailed results regarding the task success and difficulty, SUS, NASA-TLX, SSQ, and eyestrain questionnaire questionnaires can be found in Appendix H. Overall, this application achieved its goal; participants were able to follow the instructions on screen to get to their location (no participants failed any task) and most tasks were very easy to complete. However, participants were not very impressed by the application (only 3 of 9 would use it again). This was because participants felt that it did not provide any additional use compared to other GPS methods, was an additional safety hazard since they were immersed in AR rather than their real space, and many felt like they were breaking social norms by holding up a phone in public in a similar way that they would if they were "recording" other people. Heuristic evaluations were very positive and just suggested that text be edited to be more readable and make it clearer what is required to use the application (e.g., bright environment, avoid heat, etc.). A summary of usability issues found as a result of the user test is in Table 22.

Table 22

Mobile Phone: Usability Issues Identified During the Google Maps Live View User Test

Usability Issue	Description
App text and AR visuals were not in sync with one another	AR visuals would suggest to "turn now" but text stated that the participant should turn in 10 ft. This confused participants and they were unsure exactly where they needed to turn.
The app was awkward to use in public	Participants said they felt like they were impeding on bystanders' privacy and breaking social norms when holding up the phone to see the AR directions in public because it looked like they were taking videos of bystanders.
App was not integrated the real environment as much as it could be	Participants felt that the app was not as useful as it could be if it used contextual AR cues. E.g., labelling buildings nearby like "student union" and "library", visually identifying when they were on a sidewalk vs. crosswalk vs. walking path.
Usage of the app could be unsafe	Participants had a few safety concerns: AR content was opaque and could visually block real life objects. Because of this, participants were worried about running into people, cars, tripping, etc. The AR directions only described the immediate next direction. Participants could not plan their path and know it was safe by using the AR feature. For example, an incorrect path was given and asked participants to cross a busy road without a crosswalk. The researcher told participants this ahead of time to avoid the issue, and participants were concerned that they would not have known this ahead of time unless the researcher told them. They were concerned about what other dangerous situations they could have encountered. Some participants stated that it is visually obvious to bystanders that you are following directions, especially if you have the volume turned on. This was a concern because they do not want strangers with ill intent to know that they are unfamiliar with their surroundings since it could put them in an unsafe situation.

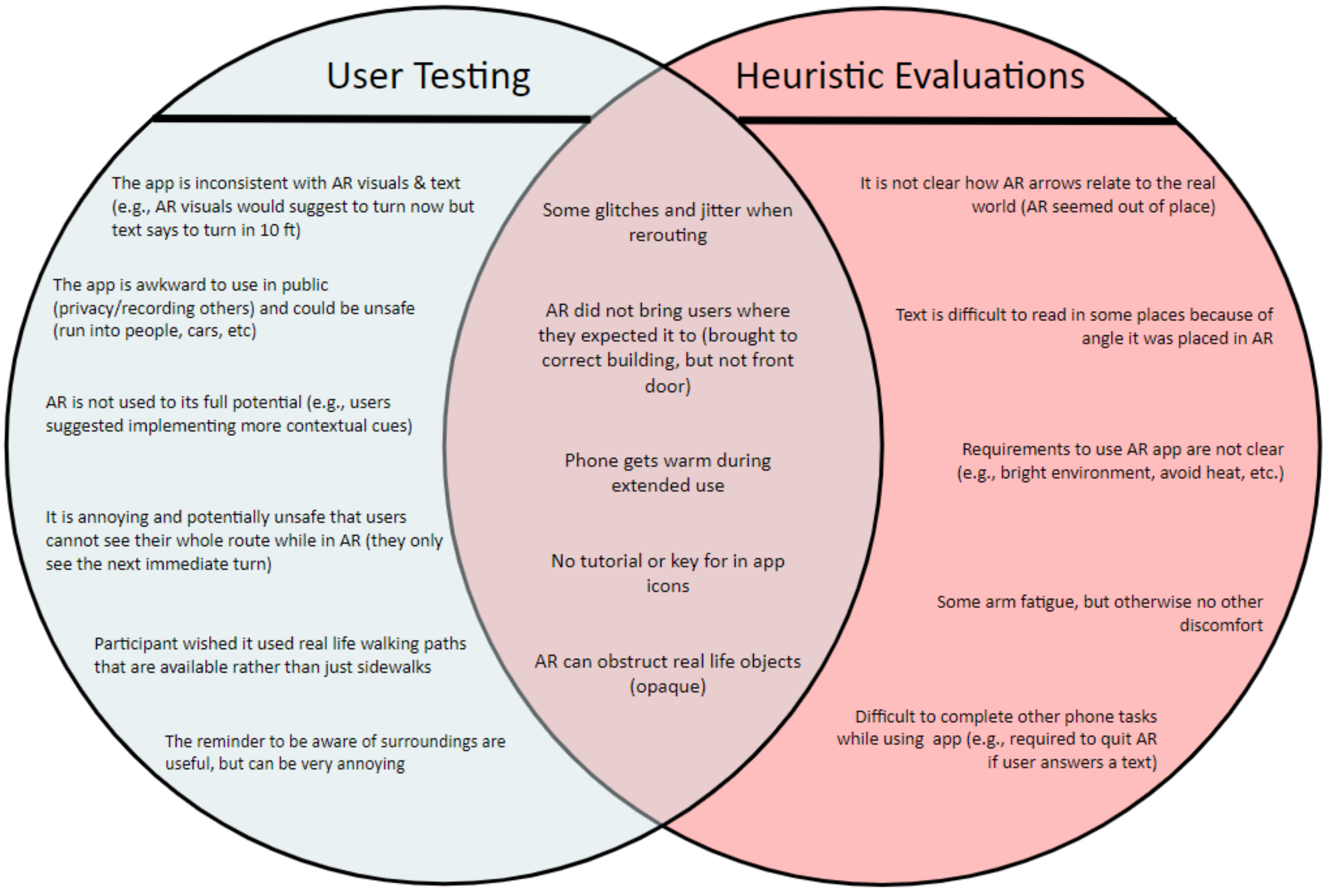
Usability Issue	Description
Unclear why feedback was being given to the user to be aware of their surroundings	The “be safe” notification was useful at points to take the participant’s eyes off their phone when in risky situations (e.g., crossing the street). But some participants found it very annoying because they did not understand what was triggering it (focusing too much on the phone, reaching a crosswalk, timed trigger, etc.)
Glitches/jitter	When rerouting, the AR content would glitch and participants had to wait a moment for it to update the AR arrows.
Did not work as expected	Participants thought that the AR directions would lead them to the door of their destination. Instead, it only took them to a side entrance to the building.
Phone became warm during extended use	Phone gets warm with extended use. This made participants not want to use this app for long walks because they expected it would become too hot to use and would drain their battery quickly.
No tutorial was provided	Participants did not encounter a tutorial, receive instruction of the meaning of AR icons, or environmental requirements to use the app. This caused ambiguity and uncertainty about how to use the app.

Discussion. Overall, the Google Maps Live View application achieved its goal; participants were able to follow the instructions on screen to get to their location (all participants completed all tasks successfully) and most tasks were very easy to complete. However, participants were not very impressed by the application (only 3 of 9 would use it again). This was because participants felt that it did not provide any additional use compared to other GPS methods, it added a safety hazard compared to other GPS methods since they were immersed in AR rather than their real space, and many felt like they were breaking social norms by holding up a phone in public in a similar way that they would if they were “recording” other people.

Results from this user test were compared to results of the heuristic evaluations of the same application. Many of the usability issues were found using both methods, especially those that involved the user interface, feedback given to the user, responsiveness of the application, glitches, instructions, and how the AR content was integrated into the real environment. Issues that were more subjective to each user's personal experience were found mostly during user testing, such as whether it was awkward to use in public and how some pop-up reminders were "annoying". It was also noted that participants had concerns about safety when using the application, it was inconsistent in its instruction, and that they wanted more AR features (e.g., text on top of buildings with building names and more information). Issues that were about long-term use (e.g., arm fatigue after extended use, and ability to complete other tasks like take phone calls while using the application), instructions and error messages, and how the AR content was integrated into the real world were mostly found in heuristic evaluations. A summary of these findings is shown in Figure 16.

Figure 166

Mobile Phone: Comparing Usability Issues Found in User Testing and Heuristic Evaluations



General Discussion of All User Tests

The user tests found many unique issues that were more subjective and based on the users' personal experiences, such as learnability issues that occurred due to participants' inexperience with AR or MR (e.g., not knowing how to move on to the next task in the application, not understanding which buttons to press on a controller, etc.), personal opinions about the functionality of the application (e.g., Google Maps Live View was "awkward" to use in public around other people, so they would not use the application again), and certain emotions they had about the application or device (e.g., that it was "cool", "unique" or if direct interactions were "fun", or "tedious").

The heuristic evaluations found more issues that were about the holistic journey of use with the device and application. These include issues such as long-term use (e.g., replacement of parts, how to report errors, reminders to take breaks, etc.), how the device impacts the privacy of users and those around them, concerns about hardware compatibility with PPE, and representation of diverse users. Most of the usability/UX issues could be found in both user tests and heuristic evaluations. However, all issues cannot be found with just one method. These two methods are not meant to replace one another, instead, they each serve their own purpose.

No additional items or examples were added, reworded, or deleted as a result of user tests. Many of the changes were already addressed by the findings from expert judgements and heuristic evaluation. Additional usability/UX issues that were found only in user tests and not heuristic evaluations were deemed too subjective or out of scope to add to the heuristic checklist. For example, comments such as, "controls were fun" were focused on subjective user satisfaction and were out of scope for the heuristic checklist.

Step 8 - Refinement Stage

Method

During step 8, all feedback from step 7 (the validation stage) is incorporated into the experimental heuristic checklist. Items or heuristics may be added, deleted, or edited as a result of this feedback. Depending on what is found during step 7, it may be necessary to repeat other previous steps (e.g., a more in-depth literature search, validation with other applications or devices, etc.). For example, during the validation stage of the Derby & Chaparro heuristic checklist (2022), it was discovered that it was necessary to further validate with more applications, devices, and investigate how to incorporate items about privacy, safety, inclusive design, collaboration, integration of physical and virtual worlds, tactile/audio features, device maintainability, and unboxing/set-up. To address this, steps 1-7 were completed, as described earlier in this chapter.

Results

Feedback gathered from the validation process resulted in the following changes that are shown in Table 23. A total of 38 changes were made to the experimental heuristic checklist that was defined in step 6 (6 items were added, 6 examples were added, 3 items were rephrased, 10 examples were rephrased, and 13 other changes were made). The most changes occurred in the Privacy, Integration of Physical & Virtual Worlds, and Consistency & Standards heuristics. One additional heuristic was created and named "Privacy".

Table 23

Changes Made to the Experimental Heuristic Checklist Based on Step 7 - Validation

Heuristic	Items Added	Examples Added	Items Reworded	Examples Reworded	Items Deleted	Other Changes
Unboxing & Set-Up	1	1	1	0	0	1 element name change
Help & Documentation	0	0	0	1	0	0
Cognitive Overload	0	0	0	1	0	0
Integration Of Physical & Virtual Worlds	3	3	0	0	0	0

Heuristic	Items Added	Examples Added	Items Reworded	Examples Reworded	Items Deleted	Other Changes
Consistency & Standards	1	1	1	2	0	0
Collaboration	0	0	1	1	0	0
Comfort	0	0	0	0	0	0
Feedback	0	0	0	1	0	0
User Interaction	0	0	0	3	0	0
Recognition Rather Than Recall	0	0	0	0	0	0
Device Maintainability	0	0	0	0	0	0
Privacy	1	1	0	1	0	1 heuristic created 5 items moved here
N/A (full checklist applies)	0	0	0	0	0	6 - moved related heuristics together; renamed heuristics; fixed auto-calculations; added instructions; changed format/theme of excel sheet; and made sure excel sheet was consistent
Total: **38**	**6**	**6**	**3**	**10**	**0**	**13**

Discussion

A total of 38 changes were made as a result of step 8. This resulted in a total of 100 changes to the Derby & Chaparro (2022) heuristic checklist as a result of step 5 and step 8 (19 additional items, 19 additional examples, 3 items rephrased, 44 examples rephrased, 2 items deleted, and 13 other changes made). The updated heuristic checklist resulted in 12 heuristics and 109 checklist items. This heuristic checklist can be found in Appendix I.

Chapter 4

Discussion

AR and MR technology has been around since the 1960s but has gained popularity in recent years. Since then, the technology and definitions have evolved. It is continuing to evolve, as we have seen with the recent announcements by Apple and Meta about their MR headsets (the Vision Pro and Quest 3) (Apple, 2023a; Meta, 2023). These companies have stated that this technology is the future of our work, communication, and entertainment. This may sound enticing, but to create usable, useful, efficient, and desirable AR and MR technology, researchers and industry professionals need to keep in mind the factors that contribute to perceived usability of their applications and devices. Many aspects related to AR and MR can impact the user's experience and how usable the device or application is. For example, how the user interacts with the device or application, how much feedback the device or application gives the user based on those interactions, how accurate the integration is between real and virtual worlds, the consistency and standards that are followed based on human perceptual capabilities, how physically comfortable and safe the device and user interactions are, if user and bystander privacy is kept a priority, how inclusive the design is, and if collaboration with other users is a feature – how well this is integrated into the application. One way to assess the usability of a device application is through heuristic evaluation. Many current AR and MR heuristics are not validated to assess any AR or MR device or application. This current study aimed to fill this gap in the research.

The purpose of this study was to create and validate a heuristic checklist that could be used to assess usability/UX issues for AR and MR applications and devices. Previous work was completed to create such a heuristic checklist (Derby & Chaparro, 2022), but it was found that the checklist needed further research and validation to meet the needs of the broad range of applications and use cases. An eight-step methodology for developing usability/UX heuristic checklists was used (Quiñones et al., 2018). Through this methodology, an experimental heuristic checklist was created based on literature found about AR and MR design (definitions, current heuristics, usability issues in the domain, results from experiments in the domain, etc.). At this point, 11 heuristics and 105 checklist items were defined. After this was created, a validation process was completed to answer the following research questions:

1. **Question:** Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more usability/UX issues** than the use of a control heuristic checklist?
2. **Question:** Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more issues that qualify as severe/critical** than the use of a control heuristic checklist?
3. **Question:** Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more domain specific issues** than the use of a control heuristic checklist?
4. **Question:** Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **similar usability/UX issues as a usability study** of the same application?

To answer these questions the following validation stages were conducted: expert reviews of the experimental heuristic checklist, heuristic evaluations that compared the results from the experimental checklist with those of a control checklist (De Paiva Guimarães & Martins, 2014), and user tests that compared the results from the experimental heuristic evaluation to user tests. Five different devices and applications were used in order to validate the experimental heuristic checklist with a broad variety of types of applications and devices. The chosen applications and devices included the following: the Meta Quest Pro (HMD) and ShapesXR, a collaborative prototyping application; the Magic Leap 1 (HMD) and Wayfair Spaces, a retail application that allows users to see virtual furniture in their own space; Magic Leap 2 (HMD) and Fan Blade Replacement, a training application for aircraft maintenance; mobile phone (iPhone 13 Pro or Android Galaxy S8) and Google Maps Live View, an AR navigation application; and Epson Moverio BT-300 (smart glasses) and a variety of applications on the device. As a result of the validation process, a total of 38 changes were made to the heuristic checklist (100 changes to the Derby & Chaparro (2022) heuristic checklist) and the four research questions were answered:

1. **Question:** Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more usability/UX issues** than the use of a control heuristic checklist?

   Answer: Yes. The expert judgement portion of step 7 (validation) confirmed that the experimental heuristic checklist covered usability/UX issues of AR and MR. Experts also provided recommendations to change the heuristic checklist to include more items regarding usability issues of AR and MR and to make the heuristic checklist easier to complete. The heuristic evaluation portion of step 7 (validation) compared results from the heuristic evaluations conducted with experimental heuristic checklist to a control heuristic checklist (de Paiva Guimarães & Martins, 2014). More usability/UX issues were found when evaluators used the experimental heuristic checklist than the control checklist for all of the 5 applications and devices that were evaluated.

2. **Question:** Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more issues that qualify as severe/critical** than the use of a control heuristic checklist?

   Answer: Yes. The expert judgement portion of step 7 (validation) identified heuristics that covered severe/critical usability issues in the experimental heuristic checklist. The heuristic evaluation portion of step 7 (validation) compared results from the heuristic evaluations conducted with experimental heuristic checklist to a control heuristic checklist (de Paiva Guimarães & Martins, 2014). More issues that qualified as severe/critical were found when evaluators used the experimental heuristic checklist than the control checklist for all of the 5 applications and devices that were evaluated.

3. **Question:** Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **more domain specific issues** than the use of a control heuristic checklist?

Answer: Yes. The expert judgement portion of step 7 (validation) identified heuristics that covered domain specific usability issues in the experimental heuristic checklist. The heuristic evaluation portion of step 7 (validation) compared results from the heuristic evaluations conducted with experimental heuristic checklist to a control heuristic checklist (de Paiva Guimarães & Martins, 2014). More domain specific issues were found when evaluators used the experimental heuristic checklist than the control checklist for all of the 5 applications and devices that were evaluated.

4. **Question:** Will the use of a heuristic checklist specifically for AR and MR devices and applications identify **similar usability/UX issues as a usability study** of the same application?

   Answer: Yes. The user testing and heuristic evaluation portion of step 7 (validation) gathered usability/UX issues for each of the 5 applications and devices. The issues found in user testing were compared to those found in the heuristic evaluations. Similar issues were found, but also unique issues were found with each method. This is to be expected because user testing and heuristic evaluations should not be a replacement for each other, as experiences between experienced heuristic evaluators and inexperienced users may differ. Most of the usability issues that were found were discovered in the heuristic evaluations, including those that related to simplicity, comfort, safety, human perceptual capabilities, glitches, collaboration, etc. Usability issues that related to learnability of the applications and devices, satisfaction and enjoyment when using the application, and immersiveness of the application were found during user testing.

The revised version of this heuristic checklist (12 items and 109 checklist items) can be found in Appendix I. Definitions of each of the new heuristics can be found in Appendix J. This heuristic checklist can help practitioners improve the usability of their applications or devices by providing a prescriptive list of aspects that impact the usability of AR and MR technology. It is encouraged that this heuristic checklist is used iteratively through the design process because it is easier to make changes to prototypes iteratively rather than at the end of the design process. A single evaluator can complete a heuristic evaluation, but since some items in this heuristic checklist are subjective and the results may change depending on what each individual evaluator encounters with the application or device, it is recommended that at least 2-5 evaluators evaluate the technology and discuss their findings together to create a summary of changes necessary to improve the usability of the technology. This is consistent with findings from others who have developed heuristic evaluations (Moran & Gordon, 2023). This heuristic checklist can also be used as a starting point for practitioners' design plans. They could look at the heuristics and checklist items for inspiration of what to keep in mind when designing their technology to enhance the usability of the product.

Limitations and future research

This study aimed to improve the Derby & Chaparro (2022) heuristic checklist by expanding its validation to different application and device types. This study expanded the validation to XR devices (Meta Quest Pro), collaborative applications (ShapesXR), mobile navigation applications (Google Maps Live View), smart glasses (Moverio BT-300), training applications (Fan Blade Replacement), and retail applications (Wayfair Spaces). However, it was not feasible to validate with every application or device type that is or ever will be in existence for AR and MR. This validation process was broad, but it is not guaranteed that this heuristic checklist will apply to every type of application or device.

Additionally, this heuristic checklist was not validated for VR applications or devices. The Meta Quest Pro is an XR device with both AR and VR capabilities, but the VR features of the ShapesXR application were not assessed in the heuristic evaluation or user testing. Usability/UX issues for VR technology will differ than AR and MR since, in VR, the user is completely immersed in a virtual environment rather than the virtual content being integrated into the user's real environment. Requirements necessary to replicate an environment in the virtual world should be addressed in a usability heuristic checklist for VR and are not addressed in this AR and MR usability heuristic checklist.

Interesting results were found when evaluating the five applications and devices. For example, participants who used the Epson Moverio BT-300 smart glasses device reported feeling more eye strain, simulator sickness symptoms, and asked for more breaks than participants who used other AR and MR devices. This was surprising since this was the most lightweight and unobtrusive HMD that was used in this current study. Future work should be conducted to understand what contributes to users' experience of cybersickness in AR and MR.

The Wayfair Spaces application included a tutorial that was very immersive, participants enjoyed it, and participants retained most of the information from it. This was very different than the tutorials provided in ShapesXR, where users reported feeling tired after watching and participants did not feel confident in their ability to use the controls of the application. These tutorials should be compared and further evaluated to understand what features of an AR or MR tutorial are important to include to make it easy, fun, and useful.

The Fan Blade Replacement application is a training application currently being developed for aircraft maintenance students at Embry-Riddle Aeronautical University. This work assessed the usability of a prototype of this application. However, work should be done to assess if this type of instruction (MR) improves retention of information when compared to paper-based training.

The Meta Quest Pro was the only HMD that used video pass-through technology to display MR information. Participants did not report any strong positive or negative experiences while using this type of MR technology. However, this may be because they were not required to view details in their real environment while using the headset (e.g., reading text on a piece of paper or on a computer screen). This should be evaluated more in depth and compared to experiences with optic see-through displays to discover how the user experiences differ depending on the technology used. Additionally, this collaborative application could be used to assess how teamwork strategies may differ in virtual space when compared to teams located in the same physical space.

Finally, participants who used the Google Maps Live View feature on mobile devices reported concerns regarding safety and privacy when using AR in public. They reported that, by holding up a phone in public, bystanders may assume that that the user is recording them and would feel uncomfortable. Additionally, participants stated they felt very immersed in the AR application, even though they were using a phone that they could easily put down, they felt that they wanted to always look at the AR content and as a result had a limited field of view. This became a safety concern when walking across streets or being around a crowd of people. More research should be conducted to investigate what features or design considerations should be implemented to make these types of mobile AR applications safer to use and designed with the users' and bystanders' privacy in mind.

Chapter 5

Conclusions

This research created and validated an AR and MR usability heuristic checklist that can be used to evaluate a broad range of AR and MR technology. This resulted in a heuristic checklist with 12 heuristics and 109 checklist items. The 12 heuristics are: Unboxing & Setting Up, Instructions, Organization & Simplification, Consistency & Flexibility, Integration of Physical & Virtual Worlds, User Interaction, Comfort, Feedback to the User, Intuitiveness of Virtual Elements, Collaboration, Privacy, and Device Maintainability. The 12 heuristics and their 109 checklist items are shown in Appendix I. The definitions of these heuristics are found in Appendix J. An Excel toolkit called, "The Derby Dozen: an AR/MR Usability Heuristic Checklist" was created to house this heuristic checklist to make evaluations easy to complete. This provides evaluators with an organized structure for checklist items, more information and examples related to the checklist items, a space for the evaluator to give a "Yes, Somewhat, No, N/A" ratings for each item, and a space to provide comments about their ratings. This is shown in Figure 17.

Figure 177

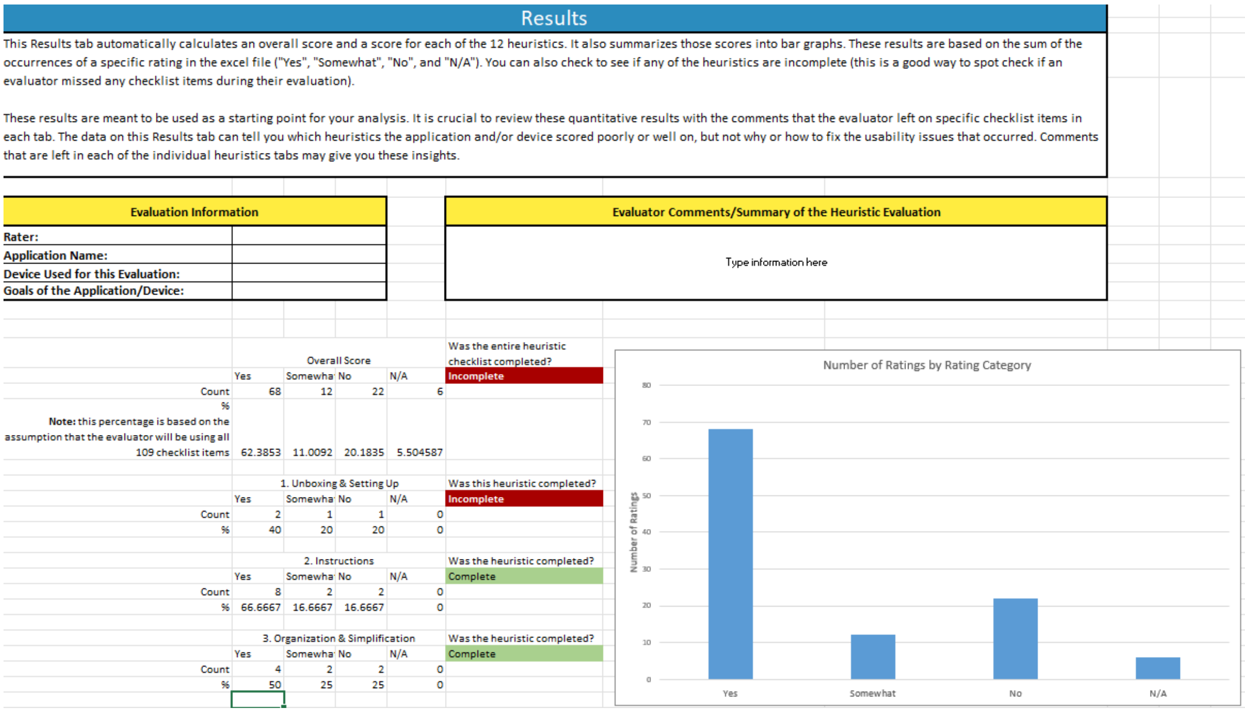
Example of a Heuristic Tab in the Derby Dozen Toolkit

Heuristic	Element	Checklist Item	More Information and Examples	Rating	Comments About Ratings
7. Comfort					
	Physiological Comfort	Can the user experience the device and/or application without pain, discomfort, nausea, disorientation, etc. DURING use?	Users should not be in pain, discomfort, or nauseous as a result of using AR/MR content. Examples of this include eye strain, skin discomfort, pain caused by the heat of the device, muscle strain, headaches, and disorientation. Allow the user to turn off any potentially hazardous functions (e.g., ensuring that flickering images are set to a minimum or can be turned off/reduced for those with epilepsy).	Yes	
		Can the user experience the device and/or application without pain, discomfort, nausea, disorientation, etc. AFTER use?	Users should not be in pain, discomfort, or nauseous as a result of using AR/MR content. Examples of this include eye strain, skin discomfort, pain caused by the heat of the device, muscle strain, headaches, and disorientation. Allow the user to turn off any potentially hazardous functions (e.g., ensuring that flickering images are set to a minimum or can be turned off/reduced for those with epilepsy).	Yes Somewhat No N/A	
		Is the device's weight light enough to feel comfortable?	Heavy devices can cause fatigue. This can also differ based on population (e.g., children may not be able to withstand the same weight of a device as an adult) or amount of physical workload as a result of gesture controls or other completing activities when using the device.	Yes	
		Does the device avoid overheating to the point that it is uncomfortable to use?	Device overheating for a head mounted display or handheld device cause the user to feel uncomfortable using the device and can become dangerous after prolonged use. It is important to avoid overheating the device to the point of discomfort for the user.	Somewhat	Overheating after extended use ~1.5 hours
		Does the device's accessories and cords avoid hindering work?	The device's cords and accessories should not get in the user's way when completing tasks.	Yes	
		Are physical interactions with the application safe and comfortable?	It is important to take into account the environment in which the device is being used. A safe gesture in a spacious environment may be safe at first, but dangerous when in an industrial setting (e.g., near large manufacturing machines)	Yes	

Additionally, this toolkit provides its users with a summary of the results found during the heuristic evaluation. This includes a data summary of the overall checklist score and individual scores of each heuristic, and graphs that both depict overall scores and scores for each heuristic. This is depicted in Figure 18. It is recommended that at least 3-5 evaluators complete a heuristic evaluation and combine their findings afterwards to find and address the most usability/UX issues as possible.

Figure 188

The Results Summary Tab in the Derby Dozen Toolkit



This current work has discussed usability and UX issues that impact the user's experience with AR and MR technology. This includes physical aspects of the device (e.g., physical comfort, safety, etc.), how the device and application function (e.g., perceptual considerations, how the real and virtual words are integrated with one another, etc.), and how the user interacts with and understands the technology (e.g., different types of user interaction, help and documentation given to the user, feedback given to the user, user privacy, collaborative interactions with others, etc.). An AR and MR heuristic checklist with 12 heuristics and 109 checklist items was created based on current literature and expert review. Through a validation process, this heuristic checklist demonstrated that it could assess such usability and UX issues in a variety of AR and MR applications. The resulting heuristic checklist can be used to assess current designs and make recommendations about how to improve the usability of the design in a way to make the experience more effective, efficient, and satisfying for users.

References

Aabel, B., & Abeywarna, D. (2018). Digital cross-channel usability heuristics: improving digital health experience. *Journal of Usability Studies, 13*(2), 52-72.

Al-Obaidi, A., & Prince, M. (2022). Usability principles for augmented reality applications in education. *IJCSNS International Journal of Computer Science and Network Security, 22*(1), 49-54.

Amburn, C. R., Vey, N. L., Mize, J. R., & Boyce, M. W. (2015). *The augmented reality sandtable (ARES)* (ARL-SR-0340). US Army Research Laboratory, Tech. https://apps.dtic.mil/sti/citations/ADA622471

Angry Birds. (2022). *Angry birds AR isle of pigs.* https://www.angrybirds.com/angry-birds-ar-isle-pigs/

Anima Res. (n.d.). *Insight Heart.* Retrieved April 29, 2022 from https://animares.com/portfolio/insight-heart

Apple (2022). *Accessibility—Foundations—Human Interface Guidelines—Design—Apple Developer.* Design. https://developer.apple.com/design/human-interface-guidelines/foundations/accessibility

Apple. (2020). *Apple AR Human Interface Guidelines.* https://developer.apple.com/design/human-interface-guidelines/augmented-reality

Apple. (2021, June 7). *Getting started with AR.* https://developer.apple.com/news/?id=c6vr1ag2

Apple. (2023a, June 5). *Introducing Apple Vision Pro: Apple's first spatial computer.* Apple Newsroom. https://www.apple.com/newsroom/2023/06/introducing-apple-vision-pro/

Apple. (2023b, May 10). *iPhone 13 Pro – Technical Specifications.* https://support.apple.com/kb/SP852?locale=en_US

Aschenbrenner, D., Leutert, F., Çençen, A., Verlinden, J., Schilling, K., Latoschik, M., Lukosch, S. (2019). Comparing human factors for augmented reality supported single-user and collaborative repair operations of industrial robots. *Front. Robot. AI, 6*(37). https://doi.org/10.3389/frobt.2019.00037

Augmented Reality Enterprise Alliance (AREA). (2021). *Augmented reality best practices safety playbook.* https://thearea.org/area-resources/augmented-reality-best-practice-safety-playbook/

Aultman, A., Dowie, S., Hamid, N. (2018). Design heuristics for mobile augmented game user interfaces. In CHI EA '18: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, 1-5. Association for Computing Machinery, New York, NY.

Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 355-385. https://doi.org/10.1162/pres.1997.6.4.355

Balani, M. S., & Tümler, J. (2021). Usability and user experience of interactions on vr-pc, Hololens 2, vr cardboard and ar smartphone in a biomedical application. In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial

Intelligence and Lecture Notes in Bioinformatics), 12770 LNCS, 275–287. https://doi.org/10.1007/978-3-030-77599-5_20

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction, 24*(6), 574-594. https://doi.org/10.1080/10447310802205776

Baños Díaz, G., & Zapata Del Río, C. M. D. P. (2018). A Proposal of usability heuristics oriented to e-banking websites. In *Design, User Experience, and Usability: Theory and Practice: 7th International Conference, DUXU 2018, Held as Part of HCI International 2018*, *Las Vegas, NV, USA, July 15-20, 2018,* 327-345. Springer International Publishing.

BBC. (2018, February 7). BBC launches augmented reality app for Civilisations. *BBC.* https://www.bbc.com/news/technology-42966371

Bekele, M. K., Champion, E., McMeekin, D. A., Rahaman, H. (2021). The Influence of Collaborative and Multi-Modal Mixed Reality: Cultural Learning in Virtual Heritage. Multimodal Technol. Interact, 5(71). https://doi.org/10.3390/mti5120079

Blokša, J. (2017). *Design guidelines for user interface for augmented reality* [Master's thesis, Masaryk University]. https://is.muni.cz/th/yombd/Thesis-final.pdf

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. *Usability evaluation in industry.*

Buker, T. J., Vincenzi, D. A., & Deaton, J. E. (2012). The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display: Reducing apparent latency with predictive compensation. Human Factors, 54(2), 235–249. https://doi.org/10.1177/0018720811428734

Burov, O., & Pinchuk, O. (2021). Extended reality in digital learning: influence, opportunities, and risks' mitigation. Educational Dimension, 5(57), 144-160.

Carter, M., & Egliston, B. (2020). Ethical implications of emerging mixed reality technologies.

Caruso, G., Polistina, S., Bordegoni, M. (2011). Collaborative mixed-reality environment to support the industrial product development. In Proceedings of the ASME 2011 World Conference on Innovative Virtual Reality, WINVR2011, Milan, Italy.

Caudell, T. P., Mizell, D. W. (1992). Augmented reality: An application of heads-up display technology to manual manufacturing processes. *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences, 2,* 659-669. https://doi.org/10.1109/HICSS.1992.183317

Chang, A., Arenas, J. J., Paz, F., & Díaz, J. (2018). Augmented reality and usability best practicies: a systematic literature mapping for educational videogames. *2018 IEEE Sciences and Humanities International Research Conference (SHIRCON).* https://doi.org/10.1109/SHIRCON.2018.8592976

Cometti, C., Païzis, C., Casteleira, A., Pons, G., & Babault, N. (2018). Effects of mixed reality head-mounted glasses during 90 minutes of mental and manual tasks on cognitive and physiological functions. PeerJ. https://doi.org/10.7717/peerj.5847

Connor, J. O., Abou-Zahra, S., Rodriguez, M. C., & Aruanno, B. (2020). XR accessibility – learning from the past and addressing real user needs for inclusive immersive environments. In: Miesenberger, K., Manduchi, R., Covarrubias Rodriguez, M., Peňáz, P. (eds) Computers Helping People with Special Needs. ICCHP 2020. *Lecture Notes in Computer Science,* (Vol. 12376), pp. 117-122. Springer, Cham. https://doi.org/10.1007/978-3-030-58796-3_15

Crets, S. (2020, October 22). *Augmented reality boosts conversion for Home Depot.* Digital Commerce 360. https://www.digitalcommerce360.com/2020/10/22/augmented-reality-boosts-conversion-for-home-depot/

De Belen, R. A. J., Nguyen, H., Filonik, D., Del Favero, D., & Bednarz, T. (2018). A systematic review of the current state of collaborative mixed reality technologies: 2013-2018. *AIMS Electronics and Electrical Engineering, 3*(2), 181-223. https://doi.org/10.3934/ElectrEng.2019.2.181

De Belen., R. A. J., & Bednarz, T. (2019). Mixed reality and internet of things (MRIoT) interface design guidelines for elderly people. *Proceedings of the 23rd International Conference in Information Visualization – Part II,* 82-85. https://doi.org/10.1109/IIV-2.2019.00025

De Guzman, J. A., Thilakarathna, K., & Seneviratne, A. (2019). Security and privacy approaches in mixed reality: A literature survey. In ACM Computing Surveys, 52(6). Association for Computing Machinery. https://doi.org/10.1145/3359626

De Paiva Guimarães, M., Martins, V. F. (2014). A checklist to evaluate augmented reality applications. *Proceedings of the 2014 XVI Symposium of Virtual and Augmented Reality,* 45-52. https://doi.org/10.1109/SVR.2014.17

Dehghani, M., Lee, S. H. M., & Mashatan, A. (2020). Touching holograms with windows mixed reality: Renovating the consumer retailing services. Technology in Society, 63, https://doi.org/10.1016/j.techsoc.2020.101394

Denning, T., Dehlawi, Z., & Kohno, T. (2014, April). In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2377-2386. https://doi.org/10.1145/2556288.2557352

Derby, J. L., & Chaparro, B. S. (2020). Use of augmented reality by older adults. In: Gao, Q., Zhou, J. (Eds.) Human Aspects of IT for the Aged Population. Technologies, Design and User Experience. HCII 2020. *Lecture Notes in Computer Science: Vol. 12207* (pp.125-134). Springer, Cham. https://doi.org/10.1007/978-3-030-50252-2_10

Derby, J. L., & Chaparro, B. S. (2022). The development and validation of an augmented and mixed reality usability checklist. In J.Y.C. Chen & G. Fragomeni (Eds.), Virtual, Augmented and Mixed Reality: Design and Development. *Lecture Notes in Computer Science: Vol. 13317.* Springer, Cham. https://doi.org/10.1007/978-3-031-05939-1_11

Desurvire, H., & Wiberg, C. (2009, July). Game Usability Heuristics (PLAY) for Evaluating and Designing Better Games: The Next Iteration. *International conference on online* *communities and social computing,* 557-566. https://doi.org/10.1007/978-3-642-02774-1_60

Dużmańska, N., Strojny, P., & Strojny, A. (2018). Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Frontiers in psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02132

Endsley, M. (2021). Situational awareness. In G. Salvendy & W. Karwowski (Eds.), *Handbook of Human Factors and Ergonomics* (5th ed., pp. 434-456). John Wiley & Sons.

Endsley, T. C., Spren, Kk. A., Brill, R. M., Ryan, K. J., Vincent, E. C., & Martin, J. M. (2017, September). Augmented reality design heuristics: Designing for dynamic interactions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*(1), 2100-2104.

Epson. (2023). *Epson Moverio BT-300 Smart Glasses.* https://www.epson.com.au/products/ProjectorAccessories/Moverio_BT-300_Specs.asp

Erickson, A., Kim, K., Bruder, G., Welch, G. (2020). A review of visual perception research in optical see-through augmented reality. *International Conference on Artificial Reality and Telexistence,* 27-35. https://doi.org/10.2312/egve.20201256

Fallahkhair, S., & Brito, C. A. (2019). Design guidelines for development of augmented reality application with mobile and wearable technologies for contextual learning. *Brazilian Journal of Technology, Communication, and Cognitive Science, 7*(1).

Franklin, F., Breyer, F., Kelner, J. (2014) Usability heuristics for collaborative augmented reality remote systems. In *Proceedings of the 16th Symposium on Virtual and Augmented Reality*, SVR, 53-62. IEEE, Piata Salvador, Brazil.

Furmanski, C., Azuma, R., & Daily, M. (2002). Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. *Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02),* 215-320. https://doi.org/10.1109/ISMAR.2002.1115091

Gabbard, J. L., Mehra, D. G., & Swan, J. E. II. (2019). Effects of ar display context switching and focal distance switching on human performance. *IEEE Transactions on Visualization and Computer Graphics, 25*(6), 2228-2241.

Gale, N., Mirza-Babaei, P., Pedersen, I. (2015). Heuristic guidelines for wearable augmented reality applications. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY*. Association for Computing Machinery, New York, NY.

Ganapathy, S. (2013). Human factors in augmented reality environments. Springer Science + Business Media, New York, NY.

Geospatial World News Desk (2016, November 8). *Trimble SketchUp Viewer for Microsoft HoloLens enables users to experience designs.* https://www.geospatialworld.net/news/trimbles-sketchup-viewer-microsoft-hololens-enables-users-experience-designs/

Goldstein, E. B. (2014). *Sensation and Perception* (9th ed.). Wadsworth, Cengage Learning.

Google. (n.d.a). *Google AR & VR*. https://arvr.google.com/ar/

Google. (n.d.b). *Glass.* Retrieved on April 29, 2022 from https://www.google.com/glass/start/

Greenfeld, A., Lugmayr, A., & Lamont, W. (2018, December). Comparative reality: Measuring user experience and emotion in immersive virtual environments. In 2*018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 204-209. IEEE.

Hale, K. S., & Stanney, K. M. (Eds.). (2014). *Handbook of virtual environments: Design, implementation, and applications.* CRC Press.

Harborth, D., & Pape, S. (2021). Investigating privacy concerns related to mobile augmented reality Apps–A vignette based online experiment*. Computers in Human Behavior, 122*. https://doi.org/10.1016/j.chb.2021.106833

Harley, A. (2018, June 3). *Visibility of system status (usability heuristic #1)*. Nielsen Norman Group. https://www.nngroup.com/articles/visibility-system-status/

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139-183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for cording data. *Communication Methods and Measures, 1*(1), 77-89. https://doi.org/10.1080/19312450709336664

Hillmann, C. (2021). *UX for XR: User experience design and strategies for immersive technologies.*

Holo-SDK (2020). *How desktop AR works.* https://www.holo-sdk.com/how-desktopar-works

Hosfelt, D., & Shadowen, N. (2020). Privacy implications of eye tracking in mixed reality. In *CHI EA '20: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI.* Association for Computing Machinery, New York, NY, United States. https://doi.org/10.1145/3334480

Hughes, C. L., Fidopiastis, C., Stanney, K. M., Bailey, P. S., & Ruiz, E. (2020). The psychometrics of cybersickness in augmented reality. *Frontiers in Virtual Reality, 1.* https://doi.org/10.3389/frvir.2020.602954

IKEA. (2022). *Say hej to IKEA place.* https://www.ikea.com/au/en/customer-service/mobile-apps/say-hej-to-ikea-place-pub1f8af050

International Organization for Standardization. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs)* (ISO Standard No. 9241-11:2018). https://www.iso.org/standard/63500.html

Jain, S., Schweiss, T., Bender, S., & Werth, D. (2021). Omnichannel retail customer experience with mixed-reality shopping assistant systems. In *Advances in Visual Computing: 16th International Symposium, ISVC 2021, Virtual Event, October 4-6, 2021, Proceedings, Part I*, 504-517. Springer International Publishing.

Jeffri, N. F. S., & Rambli, D. R. A. (2020). Guidelines for the interface design of AR systems for manual assembly. *Proceedings of the 2020 4th International Conference on Virtual and Augmented Reality Simulations,* 70-77. https://doi.org/10.1145/3385378.3385389

Joyce, A. (2020, December 13). *Help and documentation: the 10th usability heuristic*. Nielsen Norman Group. https://www.nngroup.com/articles/help-and-documentation/

Kalalahti, J. (2015). Developing usability evaluation heuristics for augmented reality applications.

Kalawsky, R. S., Stedmon, A. W., Hill, K., & Cook, C. A. (2000). A taxonomy of technology: defining augmented reality. In *Proceedings of the IEA 2000/HFES 2000 Congress.*

Kaufeld, M., Muundt, M., Forst, S., & Hecht, H. (2022). Optical see-through augmented reality can induce severe motion sickness, *Displays, 74.* https://doi.org/10.1016/j.displa.2022.102283

Kemeny, A., Chardonnet, J. R., & Colombet, F. (2020). *Getting rid of cybersickness in virtual reality, augmented reality, and simulators*. Springer Nature. https://doi.org/10.1007/978-3-030-59342-1

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203-220. https://doi.org/10.1207/s15327108ijap0303_3

Kim, S., Billinghurst, M., & Kim, K. (2020). Multimodal interfaces and communication cues for remote collaboration. In *Journal on Multimodal User Interfaces*, 14(4), 313-319.

Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s12193-020-00346-8

Kim, S., Lee, G., Huang, W., Kim, H., Woo, W., & Billinghurst, M. (2019, May). Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* 1-13*.* https://doi.org/10.1145/3290605.3300403

Ko, S. M., Chang, W. S., & Ji, Y. G. (2013). Usability principles for augmented reality applications in a smartphone environment. *International Journal of Human-Computer Interaction, 29*(8), 501-515. https://doi.org/10.1080/10447318.2012.722466

Koreng, R., & Krömker, H. (2019). Augmented reality interface: guidelines for the design of contrast ratios. *Proceedings of the ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.* https://doi.org/10.1115/DETC2019-97341

Kraub, V., & Boden, A. (2021, May 6). Current practices, challenges, and design implications for collaborative ar/vr application development. In *Proceedings of Conference on Human Factors in Computing Systems, CHI '21, Yokohama, Japan*. https://doi.org/10.1145/3411764.3445335

Krause, R. (2021, January 10). *Maintain consistency and adhere to standards (usability heuristic #4)*. Nielsen Norman Group. https://www.nngroup.com/articles/consistency-and-standards/

LaViola, J. J. (2000). A Discussion of Cyber-sickness in Virtual Environments. *SIGCHI Bull. 32*(1), 47–56. https://doi.org/10.1145/333329.333344.

Lebeck, K., Ruth, K., Kohno, T., & Roesner, F. (2018, May) Towards security and privacy for multi-user augmented reality: foundations with end users. *2018 IEEE Symposium on Security and Privacy,* 392-408*.* https://doi.org//10.1109/SP.2018.00051

Lee, N. (2019, August 8). *Spatial's collaborative AR platform is basically FaceTime in 3D. https://www.engadget.com/2018-10-24-spatial-augmented-reality-3d.html*

LEGO (2022). *Building toys with AR technology | LEGO Hidden Side.* https://www.lego.com/en-us/themes/hidden-side/about

Lenovo (2022). *ThinkReality A3.* https://www.lenovo.com/us/en/thinkrealitya3/?orgRef=https%253A%252F%252Fwww.google.com%252F

Lenovo. (n.d.). *AR, VR, MR, and Tango.* Retrieved April 9, 2022, from https://www.lenovo.com/ae/en/faqs/pc-life-faqs/ar-vr-mr-and-tango/

Leon. (2022, May 21). *Microsoft HoloLens 2 vs Magic Leap One: a comprehensive comparison.* https://vrx.vr-expert.com/microsoft-hololens-2-vs-magic-leap-one-a-comprehensive-comparison/

Leverenz, T. (2019). *The development and validation of a heuristic checklist for clinical decision support mobile applications*. [Doctoral dissertation, Wichita State University].

https://soar.wichita.edu/bitstream/handle/10057/16381/d19008_Leverenz.pdf?sequence=1&isAllowed=y

Liang, S. (2016). Design principles of augmented reality focusing on the aging population. In *Proceedings of the 30th International BCS Human Computer Interaction Conference*, pp. 1-7. BCS Learning and Development Ltd, Swindon, England.

Lindemann, P., Lee, T., & Rigoll, G. (2018). Supporting driver situational awareness for autonomous urban driving with an augmented-reality windshield display. *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct),* 358-363. https://doi.org/10.1109/ISMAR-Adjunct.2018.00104

Livingston, M. A., Gabbard, J. L., Swan, J. E., Sibley, C. M., & Barrow, J. H. (2013). Basic perception in head-worn augmented reality displays. In: Huang, W., Alem, L., Livingston, M. (eds) *Human factors in augmented reality environments*. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-4205-9_3

Lukosch, S., Lukosch, H., Dacu, D., & Cidota, M. (2015). Providing information on the spot: Using augmented reality for situational awareness in the security domain. *Comput Supported Coop Work, 24*, 613-664. https://doi.org/10.1007/s10606-015-9235-4

Magic Leap (2018). *Accessibility | Magic Leap. Magic Leap Developer*. https://ml1-developer.magicleap.com/en-us/learn/guides/bp-for-accessibility

Magic Leap. (2022). *Magic Leap homepage.* https://www.magicleap.com/en-us/

Magic Leap. (2023, September 28). *Developer-docs best practices.* https://developer-docs.magicleap.cloud/docs/category/best-practices/

Magic Leap. (n.d.a.). *Magic Leap 2.* https://www.magicleap.com/magic-leap-2

Magic Leap (n.d.b.). *Wayfair Spaces.* https://world.magicleap.com/en-us/details/com.wayfair.spaces

Marques, B., Teixeira, A., Silva, S., Alves, J., Dias, P., & Santos, B. S. (2022). A critical analysis on remote collaboration mediated by Augmented Reality: Making a case for improved characterization and evaluation of the collaborative process. *Computers & Graphics, 102,* 619-633. https://doi.org/10.1016/j.cag.2021.08.006

McGill, M. (2021). White Paper - The IEEE Global Initiative on Ethics of Extended Reality (XR) Report--Extended Reality (XR) and the Erosion of Anonymity and Privacy. In *Extended Reality (XR) and the Erosion of Anonymity and Privacy - White Paper,* 1-24.

Meta. (2019a). *Spark AR from Facebook.* https://go.facebookinc.com/SparkAR_Design_Guidelines.html

Meta. (2019b). *Designing AR experiences for everyone.* https://spark.meta.com/blog/designing-ar-experiences-for-everyone/

Meta. (2023, September 27). *Meet Meta Quest 3, our mixed reality headset starting at $499.99.* Meta Newsroom. https://about.fb.com/news/2023/09/meet-meta-quest-3-mixed-reality-headset/

Meta. (n.d.). *Meta Quest Pro Tech Specs.* https://www.meta.com/quest/quest-pro/tech-specs/

Meyers, M., Hughes, C., Fidopiastis, C., & Stanney, K. (2020). Long Duration AR Exposure and the Potential for Physiological Effects. *MODSIM World 2020.*

Microsoft. (2022). *Start Designing and Prototyping. Design Guidance for MR.* https://learn.microsoft.com/en-us/windows/mixed-reality/design/design

Microsoft. (2022, June, 7). *Learn windows mixed reality start designing and prototyping. https://developer-docs.magicleap.cloud/docs/category/best-practices/*

Microsoft. (2022a). Microsoft HoloLens | Mixed Reality Technology for Business. https://www.microsoft.com/en-us/hololens

Microsoft. (2022b). Mixed Reality docs - Start gesture. https://docs.microsoft.com/en-us/windows/mixed-reality/design/system-gesture

Microsoft. (n.d.) *Get Prism by Object Theory.* https://askus.baker.edu/faq/218130#:~:text=Company%20Name.,Retrieved%20from%20website%20address.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1994). Augmented reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and Telepresence Technologies, 2351*, 282-292. https://doi.org/10.1117/12.197321

Miller, M. R., Jun, H., Herrera, F., Villa, J. Y., Welch, G., & Bailenson, J. N. (2019). Social interaction in augmented reality. *PLoS ONE, 14*(5). https://doi.org/10.1371/journal.pone.0216290

Mine, M. R., Van Baar, J., Grundhofer, A., Rose, D., & Yang, B. (2012). Projection-based augmented reality in disney theme parks. *Computer, 45*(7), 32-40. https://doi.org//10.1109/MC.2012.154

Moran, K., & Gordon, K. (2023, June 25). *How to conduct a heuristic evaluation.* Nielsen Norman Group. https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/#:~:text=Heuristic%20evaluations%20work%20best%20when,independently%20evaluate%20the%20same%20interface

Morville, P. (2004, June 21). *User experience design.* Semantic Studios. http://semanticstudios.com/user_experience_design/

Müller, J., Rädle, R., Reiterer, H. (2017, May). Remote collaboration with mixed reality displays: how shared virtual landmarks facilitate spatial referencing. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems,* 6481-6486*.* https://dx.doi.org/10.1145/3025453.3025717

Murtza, R., Monroe, S., & Youmans, R. J. (2017). Heuristic Evaluation for Virtual Reality Systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*(1), 2067–2071. https://doi.org/10.1177/1541931213602000

Nguyen, H., & Bednarz, T. (2020). User Experience in Collaborative Extended Reality: Overview Study. *In The 17th EuroVR International Conference*, 41–70, Valencia, Spain. https://doi.org/10.1007/978-3-030-62655-6_3ï

Niantic, Pokémon, & Nintendo. (2022). *Pokémon Go live.* https://pokemongolive.com/en/

Nielsen, J. (1994a). *Enhancing the explanatory power of usability heuristics*. Proc. ACM CHI'94 Conf. (Boston, MA, April 24-28), 152-158.

Nielsen, J. (1994b). *The theory behind heuristic evaluations.* Nielsen Norman Group. https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/theory-heuristic-evaluations/

Nielsen, J. (2005). Ten Usability Heuristics. https://pdfs.semanticscholar.org/5f03/b251093aee730ab9772db2e1a8a7eb8522cb.pdf

Nielsen, J. (2012, January 3). *Usability 101: introduction to usability*. https://www.nngroup.com/articles/usability-101-introduction-to-usability/

Nielsen, J., Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 249-256. https://doi.org/10.1145/97243.97281

Nilsen, T. (2006). *Guidelines for the design of augmented reality strategy games* [Master's thesis, University of Canterbury]. UC Research Repository. https://ir.canterbury.ac.nz/handle/10092/1109

Nilsson, S. (2010). *Augmentation in the wild: user centered development and evaluation of augmented reality applications* [Unpublished doctoral dissertation]. Linköping University. http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-54241

Noah, N., Shearer, S., & Das, S. (2022, October). Security and privacy evaluation of popular augmented and virtual reality technologies. In *Proceedings of the 2022 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence, and Neural Engineering (IEEE MetroXRAINE 2022).*

O'Connor, J., Sajka, J., White, J., Hollier, S., & Cooper, M. (2021, November 4). *XR Accessibility User Requirements*. W3C. https://w3c.github.io/apa/xaur/

Ortman, E., & Swedlund, K. (2012). *Design guidelines for user interface for augmented reality* [Master's thesis, Umeå University]. http://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A558531&dswid=-1998

Pan, Y., Sinclair, D., & Mitchell, K. (2018). Empowerment and embodiment for collaborative mixed reality systems. *Computer Animation and Virtual Worlds*, *29*(3-4), e1838. https://doi.org/10.1002/cav.1838

Park, B., Hunt, S., Martin, C., Nadolski, G. J., Wood, B. J., & Gade, T. P. (2020). Augmented and mixed reality: technologies for enhancing the future of IR. *Journal of Vascular and Interventional Radiology, 31*(7), 1074-1082. https://doi.org/10.1016/j.jvir.2019.09.020

Petrovčič, A., Taipale, S., Rogelj A., & Dolničar V. (2017). Design of mobile phones for older adults: an empirical analysis of design guidelines and checklists for feature phones and smartphones. *International Journal of Human-Computer Interaction, 3*(3), 251-264. https://doi.org/10.1080/10447318.2017.1345142

Piumsomboon, T., Day, A., Ens, B., Lee, Y., Lee, G., & Billinghurst, M. (2017). Exploring enhancements for remote mixed reality collaboration. In *SA'17 Symposium on Mobile Graphics & Interactive Applications, November 27-30, 2017, Bangkok, Thailand.* https://doi.org/10.1145/3132787.3139200

Potemin, I. S., Zhdanov, A., Bogdanov, N., Zhdanov, D., Livshits, I., Wang, Y. (2018, November). Analysis of the visual perception conflicts in designing mixed reality systems. In *Optical Design and Testing VIII*: Vol 10815. https://doi.org/10.1117/12.2503397

Pratticó, F. G., & Lamberti, F. (2021). Mixed-reality robotic games: design guidelines for effective entertainment with consumer robots. *IEEE Consumer Electronics Magazine, 10*(1), 6-16. https://doi.org/10.1109/MCE.2020.2988578

Quiñones, D., Rusu, C., & Rusu, V. (2018). A methodology to develop usability/user experience heuristics. *Computer Standards & Interfaces, 59,* 109-129. https://doi.org/10.1016/j.csi.2018.03.002

Rajaram, S., Roesner, F., & Nebeling, M. (2021, January). Designing privacy-informed sharing techniques for multi-user ar experiences. In *VR4Sec: 1st International Workshop on Security for XR and XR for Security, Vancouver, B. C., Canada.*

Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). What is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior, 133,* https://doi.org/10.1016/j.chb.2022.107289

Riley, J., Flint, J., Wilson, D. P., Fidopiastis, C. M., & Stanney, K. M. (2020). Towards a Predictive Framework for AR Receptivity. In Chen, J.Y.C., Fragomeni, G. (eds) *Virtual, Augmented and Mixed Reality. Design and Interaction. HCII 2020. Lecture Notes in Computer Science, 12190*. Springer, Cham. https://doi.org/10.1007/978-3-030-49695-1_10

Rolim, C., Schmalstieg, D., Kalkofen, D., & Teichrieb, V. (2015). Design guidelines for generating augmented reality instructions. *2015 IEEE International Symposium on Mixed and Augmented Reality,* 120-123. https://doi.org/10.1109/ISMAR.2015.36

Rowen, A., Grabowski, M., Rancy, J., & Crane, A. (2019). Impacts of wearable augmented reality displays on operator performance, situational awareness, and communication in safety-critical systems. *Applied Ergonomics, 80*, 17-27. https://doi.org/10.1016/j.apergo.2019.04.013

Ruano, S., Cuevas, C., Gallego, G., & García, N. (2017). Augmented reality tool for situational awareness improvement of UAV operators. *Sensors, 17*(2), 297. https://doi.org/10.3390/s17020297

Ruth, K., Kohno, T., & Roesner, F. (2019). Secure {Multi-User} content sharing for augmented reality applications. In *the Proceedings of the 28th USENIX Security Symposium (USENIX Security 19*), 141-158, Santa Clara, CA.

Sahija, D. (2022). Critical review of mixed reality integration with medical devices for patient care. *International Journal for Innovative Research in Multidisciplinary Field, 8*(1), 100-105. https://doi.org/10.2015/IJIRMF/202201017

Salimian, M. H., Reilly, D., Brooks, S., & MacKay, B. (2016, November). Physical-digital privacy interfaces for mixed reality collaboration: an exploratory study. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*, 261-270. https://doi.org/10.1145/2992154.2992167

Santos, M., Taketomi, T., Yamamoto, G., Rodrigo, M., Sandor, C., & Kato, H. (2015). Toward guidelines for designing handheld augmented reality in learning support. *Proceedings of the 23rd International Conference on Computers in Education.*

Schneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer interaction*. Addison-Wesley.

Shukri, S., A., Arshad, H., & Abidin, R. (2017). The design guidelines of mobile augmented reality for tourism in Malaysia. *API Conference Proceedings.* https://doi.org/10.1063/1.5005359

Smith, D. (2021). *Flux VR: The development and validation of a heuristic checklist for virtual reality game design supporting immersion, presence, and flow.* [Doctoral dissertation, Wichita State University]. Shocker Open Access Repository. https://soar.wichita.edu/handle/10057/21577

Snapchat. (n.d.) *Snapchat.* Retrieved April, 29, 2022 from https://www.snapchat.com/

Stanney, K. M., Archer, J. A., Skinner, A., Horner, C., Hughes, C., Brawand, N. P., Martin, E., Sanchez, S., Moralez, L., Fidopiastis, C. M., & Perez, R. S. (2022). Performance gains from adaptive eXtended Reality training fueled by artificial intelligence. *Journal of Defense Modeling and Simulation, 19*(2), 195–218. https://doi.org/10.1177/15485129211064809

Stanney, K. M., Kennedy, R. S., & Drexler, M. (1997). Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting – 1997,* 1138-1142.

Stanney, K. M., Mollaghasemi, M., Reeves, L., Breaux, R., Graeber, D. A. (2003). Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design. *International Journal of Human-Computer Studies, 58*(4), 447-481. https://doi.org/10.1016/S1071-5819(03)00015-6

Stanney, K. M., Nye, H., Haddad, S., Hale, K. S., Padron, C. K., & Cohn, J. V. (2021). Extended reality (XR) environments. In G. Salvendy & W. Karwowski (Eds.), *Handbook of Human Factors and Ergonomics* (5th ed., pp. 434-456). John Wiley & Sons.

Stanney, K., Lawson, B. D., Rokers, B., Dennison, M., Fidopiastis, C., Stoffregen, T., Weech, S., & Fulvio, J. M. (2020). Identifying causes of and solutions for cybersickness in immersive technology: reformulation of a research and development agenda. *International Journal of Human-Computer Interaction, 36*(19), 1783–1803. https://doi.org/10.1080/10447318.2020.1828535

Sutcliffe, A., & Gault, B. (2004). Heuristic evaluation of virtual reality applications. *Interacting with Computers, 16*(4), 831–849. https://doi.org/10.1016/j.intcom.2004.05.001

Sutherland, I. E. (1968). A head-mounted three dimensional display. *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I,* 757-764. https://doi.org/10.1145/1476589.1476686

Syal, S., & Mathew, R. (2020). Threats faced by mixed reality and countermeasures. *Procedia Computer Science, 171*, 2720-2728. https://doi.org/10.1016/j.procs.2020.04.295

Tom's Guide Staff. (2018, October 3). *Samsung Galaxy S8 user guide: tips, tricks, and how-tos.* https://www.tomsguide.com/us/samsung-galaxy-s8-guide,review-4330-8.html

Tuli, N., Mantri, A. (2020). Evaluating usability of mobile-based augmented reality learning environments for early childhood. *International Journal of Human-Computer Interaction, 37*(9), 815-827. https://doi.org/10.1080/10447318.2020.1843888

Vi, S., da Silva, T., & Maurer, F. (2019). User experience guidelines for designing HMD extended reality applications. In: D. Lamas et al. (Eds.) INTERACT 2019. *Lecture Notes in Computer Science: Vol 11749* (pp. 319-341). Springer, Cham. https://doi.org/10.1007/978-3-030-29390-1_18

Vovk, A., Wild, F., Guest, W., & Kuula, T. (2018). Simulator sickness in augmented reality training using the microsoft hololens. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* 1-9. Montreal, Canada. https://doi.org/10.1145/3173574.3173783

Wang, P., Zhang, S., Bai, X., Billinghurst, M., Zhang, L., Wang, S., Han, D., Lv, H., & Yan, Y. (2019). A gesture- and head-based multimodal interaction platform for MR remote collaboration. *International Journal of Advanced Manufacturing Technology, 105*(7–8), 3031–3043. https://doi.org/10.1007/s00170-019-04434-2

Warby Parker. (2022). *Warby Parker Virtual Try-On.* https://www.warbyparker.com/app?bidkw=warby%20parker%20virtual%20try%20on&dvc=c&network=g&mobile=&searchntwk=1&content=&creative=553415437822&adposition=&placement=&target=&keyword=warby%20parker%20virtual%20try%20on&matchtype=e&mkwid=s&pmt=e&pdv=c&product_category=&utm_sour

ce=google&utm_medium=cpc&utm_term=warby%20parker%20virtual%20try%20on&utm_campaign=Brand-Glasses-Exact&cvo_campaign=Brand-Glasses-Exact&cvo_adgroup=130169137642&utm_term=130169137642-kwd-321034909347&utm_content=553415437822&singular=124__g__14963684804__130169137642__553415437822&gclid=CjwKCAjw9qiTBhBbEiwAp-GE0aWS8qmjnXxfY593vGwlnhLmhNn2cMElxI9-3Bo0eCB8eP97V9iEyRoCr6wQAvD_BwE

Wei, Z., & Landay, J. A. (2018). Evaluating speech-based smart devices using new usability heuristics. *IEEE Pervasive Computing, 17*(2), 84-96.

Wetzel, R., Blum, L., Broll, W., Oppermann, L. (2011). Designing Mobile Augmented Reality Games. In: B. Furht (Eds.), *Handbook of augmented reality*. Springer. https://doi.org/10.1007/978-1-4614-0064-6_25

Wetzel, R., McCall, R., Braun, A., Broll, W. (2008). Guidelines for designing augmented reality games. *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, 173-180. Association for Computing Machinery. https://doi.org/10.1145/1496984.1497013

White, G. R., Mirza-Babaei, P., McAllister, G., & Good, J. (2011, May). Weak inter-rater reliability in heuristic evaluation of video games. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 1441-1446.

XR Association (2018). *XR primer 2.0: a starter guide for developers*. https://xra.org/research/xr-primer-1-0-a-starter-guide-for-developers/

Yim, H. B., & Seong, P. H. (2010). Heuristic guidelines and experimental evaluation of effective augmented-reality based instructions for maintenance in nuclear power plants. *Nuclear Engineering and Design, 240,* 4096-4102. https://doi.org/10.1016/j.nucengdes.2010.08.023

Zhdanova, A. D., Zhdanov, D. D., Bogdanov, N. N., Potemin, I. S., Galaktionov, V. A., & Sorokin, M. I. (2019). Discomfort of visual perception in virtual and mixed reality systems. *Programming and Computer Software, 45*(4), 147-155. https://doi.org/10.1134/S036176881904011X

Appendix A

AR and MR Usability Heuristic Checklist (Derby & Chaparro, 2022)

This Appendix gives a brief overview of each of the 12 heuristics in Derby & Chaparro (2022). An evaluator may rate each of these checklist items as a “yes”, “somewhat”, “no”, or “not applicable (N/A)” and provide additional information about why they gave the item that rating.

Heuristic 1: Unboxing & Set-Up

1. Is the unboxing process a positive experience?
2. 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?
3. Is a quick start guide available with the device?
4. Is a call to action (QR code, instructions to use AR, etc.) clearly marked in the physical space?

Heuristic 2: Help & Documentation

5. Is there the option of a tutorial upon first use of the device and/or application?
6. Does the tutorial explain all of the necessary actions/mechanics to use the device and/or application?
7. Is the tutorial easy to understand?
8. Are required interactions easy to learn?
9. Is help or documentation easily accessible for the application?
10. Are instructions easy to understand?

11. Do instructions provide actionable feedback?

12. If auditory instructions are given, do these instructions match what the user is seeing in the application?

13. Are error messages easy to understand?

14. Do error messages provide actionable feedback?

15. Does the device's user interface and/or application avoid irreversible errors?

16. Is there a way for the user to report errors or crashes to the developer?

Heuristic 3: Cognitive Overload

17. Is the user eased into the virtual environment?

18. Does the device's user interface and/or application avoid clutter, as appropriate?

19. Does the device's user interface and/or application avoid large amounts of text?

20. Does the screen space focus on the virtual elements rather than controls or other non-AR features, as appropriate?

21. Is information organized in an understandable manner?

22. If the quantity of information is large, is it organized in a layered or hierarchical manner so it is easy to understand?

23. Does the device's user interface and/or application avoid tasks that involve a large amount of steps to complete?

24. Does the application make use of all of its AR functions (including information that is visual, auditory, and involved other sensory modalities)?

Heuristic 4: Integration of Physical & Virtual Worlds

25. Are physical (real-world) elements easily distinguishable from virtual elements?

26. Is it clear which virtual elements can be interacted with and which cannot?

27. Does the device's user interface and/or application avoid obstructing physical or virtual elements that are necessary for the users' goals?

28. Does the device's user interface and/or application avoid obstructing virtual navigation elements?

Heuristic 5: Consistency & Standards

29. Are virtual elements easy to delete or close out of?

30. Can the user pause the application at any point?

31. Are all aspects of the device's user interface and/or application (virtual elements, controls, text, etc.) clear and readable?

32. Are virtual elements sized appropriately?

33. Are virtual elements rendered a reasonable distance away from the user's targeted point?

34. For mobile devices, are the controls based on known interactions for mobile devices?

35. For mobile devices, are landscape and portrait mode supported?

36. For mobile devices, is the application responsive?

37. Do virtual elements act as the user would expect them to in the real world?

38. Are virtual elements accurately placed on the real environment?

39. Does the device and/or application avoid lag, delays, jitter, drift, and other forms of virtual element malfunctions?

40. Is the navigation consistent throughout the device and/or application?

41. Can the user navigate freely throughout aspects of the device and/or application?

42. Does the device adjust to the environment it is used in?

43. Are environmental requirements clearly defined?

44. Does the device and/or application remind users to be aware of their surroundings?

45. Does the device and/or application avoid jargon?

46. Are sans serif font types used, as appropriate, throughout the device and/or application?

47. Is the contrast between the background and text sufficient enough that the text can be read easily under a range of normal lighting conditions?

48. If the text background is transparent, is the text visible across different backgrounds and under a range of normal lighting conditions?

49. Is the volume adjustable so the user can hear audio, even in noisy environments?

50. Are auditory features understandable?

51. Are captions available for auditory features as appropriate?

Heuristic 6: Collaboration

52. Does the device and/or application allow for the user to control privacy-related content?

53. Is it clear what information is private or public content?

54. Does the application allow users to preserve virtual elements from others users' changes?

55. Is it clear which virtual elements can be interacted with and which cannot for each person?

56. Is content consistent across all users, as it is appropriate?

Heuristic 7: Comfort

57. Can the user experience the device and/or application without pain, discomfort, nausea, etc. DURING use?

58. Can the user experience the device and/or application without pain, discomfort, nausea, etc. AFTER use?

59. Is the device's weight light enough to feel comfortable?

60. Does the device avoid overheating to the point that it is uncomfortable to use?

61. Are physical interactions with the application safe and comfortable?

62. Does the application avoid making the user walk backwards, pull their head back, or push their head downwards to see virtual elements?

63. Do interactions with the device and/or application avoid tiring the user?

64. Does the device and/or application avoid causing the user eye strain?

65. Are users reminded to take breaks to prevent eye strain and fatigue?

66. Does the device easily adjust its size for different users?

67. Does the device accommodate users with eyeglasses?

68. Does the device accommodate for personal protective equipment?

Heuristic 8: Feedback

69. Does the device and/or application provide feedback on its status?

70. Does the device and its accessories provide feedback about battery levels and charging state?

71. Does the device and/or application provide feedback for user input?

72. Does the device and/or application respond quickly to user input?

73. Does the device and/or application provide the user feedback after automatic selections?

74. If an automatic selection occurs, does the device and/or application suggest what to do next?

Heuristic 9: User Interaction

75. Does the user feel in control?

76. Are user interactions simple and easy to understand?

77. Does the device and/or application include multiple forms of interaction so users can choose based on ability, preference, & skill?

78. Are the forms of interaction direct when it is appropriate to use this form of interaction?

79. Does the device and/or application avoid interactions that force the user to make large or sudden movements?

80. Does the device and/or application accommodate for the user to complete other necessary real-world tasks?

81. Does object manipulation work well in all instances?

82. Do virtual elements adapt to the users' position appropriately?

83. Does the device and/or application avoid input overloading by assigning distinct functions to buttons or gestures?

Heuristic 10: Recognition Rather than Recall

84. Are virtual elements and icons self-explanatory (does their form communicate function)?
85. Are virtual elements and controls placed near objects they reference?
86. If a virtual element is related to an object that is in motion, is the virtual element tightly coupled with object in motion appropriately?
87. Are virtual elements that are outside of the field of view easy to find?
88. Are available user actions identifiable?
89. If voice commands are included, are text labels for voice commands given?

Heuristic 11: Device Maintainability

90. Is the device sturdy enough to withstand multiple uses?
91. Does the device have a sturdy storage case?
92. Is it easy to clean the lenses, cameras, and other components on the device?
93. Are device parts fixable and replaceable as needed?
94. Does the device's battery life last long enough to perform necessary tasks of the application?

Appendix B

Summary of Usability Studies and Review Articles Found as a Result of Steps 1 and 2

Citation	Area of Focus	Type of Article
Noah et al., 2022	AR and VR User Privacy	Review
Marques et al., 2022	AR Collaborative Apps	Usability Study
Miller et al., 2019	AR Collaborative Apps	Usability Study
Kalawsky et al., 2000	AR Definitions	Review
Gabbard et al., 2019	AR Physiological Risks	Usability Study
Stanney et al., 2021	AR Physiological Risks	Usability Study
Hughes et al., 2020	AR Physiological Risks (Cybersickness)	Usability Study
Kaufeld et al., 2022	AR Physiological Risks (Cybersickness)	Usability Study
Meyers et al., 2020	AR Physiological Risks (Cybersickness)	Review
Vovk et al., 2018	AR Physiological Risks (Cybersickness)	Usability Study
Riley et al., 2020	AR Receptivity Framework	Theoretical Model
Denning et al., 2014	AR User Privacy	User Study
Ruth et al., 2019	AR User Privacy	User Study
Lebeck et al., 2018	AR User Privacy for Collaborative Apps	User Study
Harborth & Pape, 2021	AR User Privacy for Mobile Apps	User Study
Bekele et al., 2021	MR Collaborative Apps	Usability Study
Caruso et al., 2011	MR Collaborative Apps	Usability Study
Müller et al., 2017	MR Collaborative Apps	Usability Study

Citation	Area of Focus	Type of Article
Piumsomboon et al., 2017	MR Collaborative Apps	Usability Study
Wang et al., 2019	MR Collaborative Apps	Usability Study
Cometti et al., 2018	MR Cybersickness	Psychological Study
Carter & Egliston, 2020	MR Ethics	Review
Dehghani et al., 2020	MR Ethics	Review
Syal & Mathew, 2020	MR Ethics	Review
Jain et al., 2021	MR User Experience	User Study
De Guzman et al., 2019	MR User Privacy	Review
Hosfelt & Shadowen, 2020	MR User Privacy	Review
Salimian et al., 2016	MR User Privacy for Collaborative Apps	User Study
Sahija, 2022	MR User Privacy for Medical Devices	Review
Rajaram et al., 2021	XR Collaboration and User Privacy	User Study
Greenfield et al., 2018	XR Collaborative Apps	Usability Study
Kim et al., 2020	XR Collaborative Apps	Usability Study
Kraub et al., 2021	XR Collaborative Apps	Usability Study
Nguyen & Bednar, 2020	XR Collaborative Apps	Review
Kemeny et al., 2020	XR Cybersickness	Review
Kennedy et al., 1993	XR Cybersickness	Scale Development
Stanney et al., 1997	XR Cybersickness	Review

Citation	Area of Focus	Type of Article
Stanney et al., 2020	XR Cybersickness	Review
Buker et al., 2012	XR Physiological Risks (Cybersickness)	User Study
Burov & Pinchuk, 2021	XR Physiological Risks (Cybersickness)	Research Article
Stanney et al., 2022	XR Training	Case Study
Balani et al., 2021	XR User Experience	Usability Study

Appendix C

Table Results from Step 3

Topic	Collected Information	Selected Information	Priority
Information about AR & MR	The most common definition of AR comes from Azuma (1997) Azuma states that the virtual content can involve "2D graphics, 3D graphics, sound, and video". Milgram & Kishino's (1994) Virtuality Continuum graphic is also often shown to describe AR. Common forms of AR devices are: mobile (tablet or smartphone), head-mounted (google glass, HoloLens), marker-based, or projection-based (similar to a HUD) - updated Milgram & Kishino's Virtuality Continuum graphic by Stanney et al., 2021 - Lenovo's definitions clarifying the distinction between AR & MR (n.d.) (simply overlaying info onto the real world vs. combing it with computer vision) Types of AR and MR include (in order of popularity) - Mobile (phone/tablet) - Wearable AR Devices such as HMDs or smart glasses (Optic see through vs. Video see through) - projection-based AR/MR - Desktop AR All of the types of AR and MR differ on function by being either marker-based, marker-less, or location-based	Augmented Reality is a technology that overlays 3-D virtual elements onto the real world. The user can interact with these virtual elements in real time. AR applications are often viewed through smartphones and head-mounted displays (hardware) and can be maker-based, marker-less, or location-based (software) . Mixed Reality differs from this as it merges the virtual information with the real environment rather than overlays it. The user can interact with these virtual elements in real time. MR applications are often viewed through smartphones and head-mounted displays (hardware) and can be maker-based, marker-less, or location-based (software) .	(3) Highly important

Topic	Collected Information	Selected Information	Priority
Features specific to AR & MR	General features of AR/MR: User-Information (enjoyment, visibility, etc.) ; User-Cognitive (consistency, learnability, predictability, etc.); User-Support (error management, help & documentation, etc.); User-Interaction (feedback, responsiveness, low physical effort, etc.); User-Usage (navigation, availability, exiting, etc.) (Ko et al., 2013) Specific features of AR/MR: Content accuracy (info about system AND/OR lag/jitter etc.); privacy of content (for collaborative spaces); virtual/environmental distinctions; energy usage/battery life; safety (cybersickness, eye strain, heat issues, safety of accessories like tethers, etc.); privacy of user data; situational awareness	Both general and specific features will be kept and used for the heuristic checklist.	(3) Highly important
Usability & UX Attributes	- ISO standard: effectiveness, efficiency, and satisfaction - Usability attributes (Nielsen): learnability, efficiency, memorability, errors, satisfaction - UX (Morville): useful, usable, desirable, findable, credible, accessible, valuable - Usable design (Usability BOK - principles for usable design): Usefulness, consistency, simplicity, communication, error prevention and handling, efficiency, workload reduction, usability judgment	Nielsen's, Morville, and the Usability BOK will be combined and used for Usability and UX attributes. Each are more comprehensive than the ISO standard for usability and while they have some overlap, each have unique components.	(2) Somewhat important

Topic	Collected Information	Selected Information	Priority
Existing Sets of Heuristics & Guidelines	Aultman et al (2018) - GUI focused Endsley et al (2017) Santos et al (2016) - Mobile Gale et al (2015) - Wearable Kalalahti (2015) de Paiva Guimarães & Martins (2014) Franklin (2014) - Collaboration focused Ko et al (2013) – Mobile Ganapathy (2013) - Mobile XR Association (2018) Liang (2018) - Mobile & Older Adult Focus O'Brion (2014) - Enterprise Apple AR Human Interface Guidelines (2020) Design Guidance for MR, Microsoft (2022) Magic Leap Design Principles (2019) AREA AR Best Practices Safety Playbook (2021) - Safety AR/MR Stanney et al. (2003) - VE W3C XR Accessibility User Recommendations (2021) ML1 Developer Kit for Accessibility (2022) Spark AR designing AR Experiences Meta (2019) Spark AR developer kit Meta (2019) Apple Designing for Accessibility (2022) - NOT AR/MR specific IEEE McGill (2021) Safety Report	de Paiva Guimarães & Martins heuristics (2014) were chosen, even though they were developed with marker-based applications in mind. These heuristics were the only heuristic set with a checklist that encompassed the most amount of features specific to AR and known usability problems.	(2) Somewhat important

Known Usability Problems	Accuracy of tracking tech (especially with environmental differences such as lighting); Lack of standards; Readability of text; Learnability of user controls; Reliability of controls; Terminology; Layout and navigation; accessibility; cognitive considerations (e.g., cognitive overload due to overload of information); flexibility and user control - eye strain can cause fatigue, discomfort, & cybersickness - cybersickness (physical, nausea, oculomotor discomfort, and disorientation) can decrease comfort and lessen the experience with the device/app (could be due to tech/display issues like glitches) - privacy can lead to intent to use or not use based on how app collects/protects/uses data and how that info is given to the user, and in collab environments (who can see my info) as well as bystanders seeing the device in public, giving agency to user about what the device collects, etc.) - safety can negatively impact the experience (tethers limiting mobility, hot device, - collab - people need to understand expectations for what they can / cannot do / see in collab environment and what the other person can do/see - inclusion, diversity & accessibility - not designing for diverse population can cause someone to have a limited experience or negative experience with the application/device (think about users with disabilities as well as a diverse set of user differences when building devices/apps)	Many of these usability issues overlap with already identified heuristics. These usability and UX attributes will be taken into consideration when creating the new checklist.	(3) Highly important

Appendix D

Control AR Heuristic Checklist from de Paiva Guimarães & Martins (2014)

Some of the language used in the items were edited for this study so they could generalize across different technologies. For example, item 14, "Is it easy to remember the application's functionalities? (i.e., is it easy to memorize the functionalities of each button or gesture?) was changed from, "Is it easy to remember the application's functionalities? (i.e., is it easy to memorize the functionalities of each marker?)"

Heuristic 1: Visibility of System Status

1. Do you know what is going on during all of the interactions?
2. Is the loading time of virtual objects in the scene satisfactory?

Heuristic 2: User Control & Freedom

3. If the application detects more than one interactable virtual object in the scene, is it possible to specify one?
4. Is it possible to execute "redo" or "undo" easily? (E.g., can the user return to a previous state without the virtual object?)

Heuristic 3: Satisfaction

5. Does the application achieve the goal?
6. Are you satisfied with the interaction solution?
7. Are you satisfied with the freedom to move around during interactions? (E.g., you don't need to look at the camera constantly.)

Heuristic 4: Aesthetic and Minimalist Design

8. Is the number of virtual objects in the scene appropriate?
9. Is the number of interaction options satisfactory? (These include markers, keyboard, mouse, joystick, etc.)

Heuristic 5: Match Between System & Real World

10. Are the virtual objects merged correctly with the real world? (This includes their position, texture, and scale)
11. Is the virtual object animation coherent with the real world?

Heuristic 6: Consistency & Standards

12. Are actions/feedback standardized? (E.g., boarders are added to the outside of all interactable objects.)

Heuristic 7: Error Prevention

13. Is error prevention enabled? (I.e., if the user selects an invalid object, is an error message presented to the user?)

Heuristic 8: Recognition Rather than Recall

14. Is it easy to remember the application's functionalities? (I.e., is it easy to memorize the functionalities of each button or gesture?)

Heuristic 9: Flexibility & Efficiency of Use

15. Is the learning curve at an acceptable level for novice users?
16. Can expert users utilize the application in an optimized manner? (E.g., can they skip introductory videos?)
17. Is it easy to adjust the device to an appropriate position and orientation to detect the user's space?

Heuristic 10: Help Users Recognize, Diagnose, and Recover from Errors

18. Is the user instructed about what to do during the interaction? (This includes instructions like, "use a gesture to open a hologram", or is there a manual?)

Heuristic 11: Environment Configuration

19. Are there specific requirements? (E.g., camera, marker, mobile, GPS, user position, lighting, print, calibration, etc.?)

Heuristic 12: Accuracy

20. Is the tracker system stable?
21. If the tracker system detects more than one object/marker in a scene, does the application continue to function correctly?

Appendix E

Definitions for the Experimental Heuristic Checklist Compiled as a Result of Step 6

<table>
<tr><td>ID</td><td>H1</td></tr>
<tr><td>Priority</td><td>(2) Important</td></tr>
<tr><td>Name</td><td>Unboxing & Set-Up</td></tr>
<tr><td>Definition</td><td>Getting started with the AR/MR device/application should be easy to identify, complete, and a positive experience.</td></tr>
<tr><td>Explanation</td><td>A user’s first experience is very important because it sets the tone and future experience with the device/application. If the set-up process is difficult, it could deter the user from any future use with the device.</td></tr>
<tr><td>Application feature</td><td>User-Information; enjoyment; familiarity; learnability</td></tr>
<tr><td>Examples</td><td>When the user takes the device out of its packaging, it should be clear what each of the pieces do and how to assemble them (if necessary). Unboxing the device should be easy to do, the user should not have to struggle with taking the device out of its packaging. For a mobile experience, it should be clear how to interact with the AR/MR content (e.g., a call to action like a QR code or other identifying information).</td></tr>
<tr><td>Benefits</td><td>An easy unboxing experience can create a seamless and positive first experience. This can directly affect the user’s first impressions with the device, setting a positive tone for future interactions.</td></tr>
<tr><td>Problems</td><td>Misunderstanding of this heuristic can cause the user to feel confused or frustrated and can lead to abandonment of use. If this unboxing and set-up process is difficult, there is potential that the device could be accidentally damaged.</td></tr>
<tr><td>Checklist Items</td><td>
<table>
<tr><th>Element</th><th>Checklist item</th><th>More information and Examples</th></tr>
<tr><td>Device Unboxing (first-time usage)</td><td>Is the unboxing process a positive experience?</td><td>This can be both physical and emotional. The physical unboxing of the device itself is easy, understandable, and/or does not harm the device. Emotionally, unboxing is exciting or enjoyable for the user.</td></tr>
<tr><td>Device Set-Up & Configuration (first-time usage)</td><td>When the user interacts with the device for the first time, are they introduced to the user</td><td>This can be in the form of a welcome page, initial introduction, tutorial, etc.</td></tr>
</table>
</td></tr>
</table>

			interface, basic interaction methods, and basic features/content?	
			Is a quick start guide available with the device?	A quick start guide is a brief instruction manual that describes how to begin using the AR device and risks to keep in mind when using the device (e.g., be aware of your environment, take breaks, etc.). This is often provided in the original box with the device, but can also be provided online. It is often more brief than a tutorial, only providing a simplified overview of how to start using the device.
		Application Set-Up	Is a call to action (QR code, instructions to use AR, etc.) clearly marked in the physical space?	This is often done for virtual elements that are pinned to a specific location (E.g., If the only way to see an AR message is to take a picture of a real-world sign with the app). This often applies to marker-based applications, and may not apply to all applications.
Usability attributes	Satisfaction; desirable			
UX factors	Simplicity; efficiency			
Set of heuristics related	Santos, 2016; de Paiva Guimarães & Martins, 2014			

<table>
<tr><td>ID</td><td>H2</td></tr>
<tr><td>Priority</td><td>(3) Critical</td></tr>
<tr><td>Name</td><td>Help & Documentation</td></tr>
<tr><td>Definition</td><td>Help and documentation for the app and device should be easily accessible and easy to understand (including tutorials). Instructions and error messages should give users clear feedback.</td></tr>
<tr><td>Explanation</td><td>Help, documentation, instructions, and error messages are necessary after an issue has occurred. Users should be able to easily find, understand, and address all information that will help them resolve the issue quickly and easily.</td></tr>
<tr><td>Application feature</td><td>User-Support; error management; help & documentation</td></tr>
<tr><td>Examples</td><td>An informative and easy to follow tutorial should be available when the user uses the device/application for the first time. This tutorial and other documentation should be available and easy to access during subsequent uses.
Error messages should include actionable items to help the user recover from the errors such as, "move closer,” "aim the device towards a flat surface," or “move to an area with more light”</td></tr>
<tr><td>Benefits</td><td>Even if a device/application is designed to be simple and intuitive, it is important that users have all documentation that can help them solve a problem if it occurs. This will increase the user’s understanding of the device/application and help them complete all of the necessary tasks.</td></tr>
<tr><td>Problems</td><td>Misunderstanding this heuristic could lead users to make guesses about how to use the device/application. This could then lead to frustration and abandonment of use if they cannot locate information about questions they have.</td></tr>
<tr><td>Checklist Items</td><td>
<table>
<tr><th>Element</th><th>Checklist item</th><th>More information and Examples</th></tr>
<tr><td>Tutorial</td><td>Is there the option of a tutorial upon first use of the device and/or application?</td><td>This may be called something else, such as "introduction" or "start here". These are an interactive series of instructions about how to begin using a device or application.</td></tr>
</table>
</td></tr>
</table>

			Does the tutorial explain all of the necessary actions/mechanics to use the device and/or application?	The tutorial should explain how to interact with the device (through gestures, voice, button controls, etc.), the meaning of menu options, necessary interactions (e.g., letting the user know that they may be required to walk around the environment; points out where to find more information if the user needs help; etc.), risks to keep in mind when using the device (e.g., be aware of your environment, take breaks, etc.), and how they should proceed (e.g., hold device near an engine to see more info, shoot at flying enemies, find hidden gems, click on options to learn more about the human heart, etc.)
			Is the tutorial easy to understand for both novice and experienced users?	The tutorial should explain the device and/or application well and in a simple manner. Novice users may need more information to get started than experienced users. One way this could be addressed is allowing experienced users to easily skip over tutorial information as appropriate.

		Are required interactions easy to learn?	The actual interactions with the device should be easy to learn and **not** complicated.
	Instructions	Is help or documentation easily accessible for the application?	E.g., a list of dictation and/or gesture commands, on-hover tool tips, or a link to customer service/online help.
		Are instructions easy to understand?	Instructions should be easy to understand and follow. The user should **not** be confused about what the instructions are telling them to do.
		Do instructions provide actionable feedback?	E.g., "move closer" or "aim the device towards a flat surface".
		If auditory instructions are given, do these instructions match what the user is seeing in the application?	The user can become confused if auditory instructions are not aligned with what they are seeing on the virtual screen. As the user interface or application is updated, the auditory instructions should also be updated to match. E.g., if the auditory instructions say "select the hand icon" the user should see a hand icon on the screen.
	Error Messages	Are error messages easy to understand?	Error messages should be easy to understand and follow. The user should **not** be confused about what the error messages are notifying them about.

		Do error messages provide actionable feedback?	E.g., "move closer" or "aim the device towards a flat surface".
		Does the device's user interface and/or application avoid irreversible errors?	E.g., includes a back button.
		Is there a way for the user to report errors or crashes to the developer?	E.g., includes a report crash button, a contact button, email address, etc.
Usability attributes	Effectiveness; satisfaction; errors		
UX factors	Usefulness; error prevention & handling		
Set of heuristics related	Ko, 2013; de Paiva Guimarães & Martins, 2014		

ID	**H3**		
Priority	(2) Important		
Name	Cognitive Overload		
Definition	The application should minimize cognitive overload by easing the user into the environment and avoiding unnecessary clutter.		
Explanation	Reducing clutter decreases the amount of cognitive overload a user experiences.		
Application feature	User-Cognitive; consistency; learnability; predictability; recognition; user-information; hierarchy		
Examples	There should be a clear layout with no distracting elements (e.g., large blocks of text, many different types of AR/MR elements obstructing the view of important physical objects that may be important for a task, etc.). The user should also be eased into the virtual environment (e.g., NOT telling a user to dodge projectiles as soon as they enter the game).		
Benefits	When devices/applications have a minimalistic and aesthetically pleasing design, users are able to find information quickly and easily.		
Problems	Misunderstanding this heuristic could potentially lead to a poor user experience with the application due to visual clutter and potential confusion of the important elements with irrelevant information.		
Checklist Items	**Element**	**Checklist item**	**More information and Examples**
		Is the user eased into the virtual environment?	For example, an application would **NOT** be easing a user in if they were told to

			dodge projectiles as soon as they entered the game
	Clutter	Does the device's user interface and/or application avoid clutter, as appropriate?	This can include virtual elements, icons, text, buttons, etc. Clutter should be avoided without sacrificing clarity of how to interact with the application. One way this can be done is by only showing controls when applicable (e.g., showing "trash" icon only when a user has selected an object)
		Does the device's user interface and/or application avoid large amounts of text?	Large blocks of text can be distracting, obstruct real-world environment or hazards, cause eye strain, or fatiguing to read. Text should be broken up into small and simple chunks wherever possible.
		Does the screen space focus on the virtual elements rather than controls or other non-AR/MR features, as appropriate?	In many cases, the virtual elements should be the focus of the screen rather than controls or other AR/MR features. However, this can depend on device and the task at hand (such as some navigational tasks).
	Organization	Is information organized in an understandable manner?	It should make sense where information (e.g., virtual elements, text, or aspects of the user interface) reside in the application. Users should not have

			to search around the application to find what they need, as this can become frustrating.
		If the quantity of information is large, is it organized in a layered or hierarchical manner so it is easy to understand?	This is often seen with menus, lists, or by filtering different types of information given through AR/MR (e.g., selecting "electrical wiring" or "pluming pipes" for a house instead of only giving the user the option to see all of the information at once). This layering of information can minimize clutter and create organization.
		Does the device's user interface and/or application avoid tasks that involve a large amount of steps to complete?	Tasks that involve a large amount of steps to complete are difficult to remember how to do, and can frustrate the user. E.g., to view more details of a virtual brain the user has to select the object, click on the menu option, select the specific highlighted area of the brain they want to see more info on, select the type of information they want to see, and filter out irrelevant info.
		Does the application make use of all of its AR/MR functions (including information that is visual, auditory,	All functions in an AR/MR application should have relevance or a use.

		and involved other sensory modalities)?	

Usability attributes	Efficiency; learnability; memorability; satisfaction; desirable
UX factors	Desirable, findable; useful; consistency; workload reduction; simplicity; efficiency
Set of heuristics related	Nielsen's 10 (Nielsen, 1994a); de Paiva Guimarães & Martins, 2014; Ko, 2013; Liang, 2018

ID	**H4**
Priority	(2) Important
Name	Integration of Physical & Virtual Worlds
Definition	It should be easy to identify virtual elements and which virtual elements are intractable. Virtual elements should not obstruct physical objects in the users' environment that are crucial for the completion of their goals.
Explanation	When there are multiple types of virtual elements in an AR/MR space (e.g., menu items, descriptions, holograms, etc.) it should be clear exactly which elements are able to be interacted with and which are only informative or are present for aesthetic. These virtual elements should flow well with the physical world. They should not obstruct physical objects that are safety risks (e.g., cords on the ground) or important for the task at hand.
Application feature	User-interaction; low physical effort; feedback; responsiveness; direct manipulation
Examples	Virtual elements should not block menu options (or other virtual navigational elements), potential risks in the physical environment (e.g., cords on the ground or low hanging materials), or physical objects that are important for the user's goals (e.g., machinery, an instruction pamphlet, etc.)
Benefits	A seamless integration of physical & virtual worlds can help the user learn how to interact with the virtual content and can increase the user's immersion with the device.
Problems	Misunderstanding of this heuristic could cause the user to be confused and unsure about how to interact with the virtual elements. A risk to user safety may occur if the virtual elements obstruct physical hazards.
Checklist Items	

Element	**Checklist item**	**More information and Examples**
	Are physical (real-world) elements easily distinguishable from virtual elements?	The user should be able to easily understand if an object they want to interact with is virtual or real.

			These can be user interface elements, AR/MR virtual elements, auditory elements, etc.
		Is it clear which virtual elements can be interacted with and which cannot?	Some virtual elements may only be included for aesthetics and cannot be interacted with (e.g., a sea background behind an interactable fish hologram). It should be clear which virtual elements can be interacted with and which cannot.
		Are virtual elements accurately placed on the real environment?	The system should know what objects need to be recognized. E.g., It can detect names of real-life equipment parts scanned and shows the correct placement of a new part. Virtual elements that are associated with a specific real-world object should accurately correspond with that object.
		Is it clear to both users and bystanders when captures and/or recordings are being taken?	E.g., a light on the user's screen and a light on the device itself, audio stating that a capture is being taken, etc.
		Is it clear how virtual objects relate to the real world environment?	E.g., a shadow on a table showing where the object will land if it is placed, shadows and highlights on virtual objects changing as they move due to the

				light source changing positions, etc.
		Obstruction	Does the device's user interface and/or application avoid obstructing physical or virtual elements that are necessary for the users' goals?	The device and/or application should always allow the user to see physical or virtual elements that are necessary for them to complete their task or goals. This includes real and virtual objects, people, controls, and text. Obstruction such as virtual instructions that obstruct the necessary real-world placement of an object, holograms that block virtual controls, virtual elements that block important text, etc. should be avoided.
			Does the device's user interface and/or application avoid obstructing virtual navigation elements?	It should be easy to navigate throughout the application. AR/MR or non-AR/MR virtual elements should not block menu options or back buttons.
Usability attributes	Effectiveness, efficiency; satisfaction			
UX factors	Findable; usable; communication; accessible			
Set of heuristics related	Kalalahti, 2015; Liang, 2018; Aultman, 2018; Ganapathy, 2013; Santos, 2016			

ID	**H5**
Priority	(2) Important
Name	Consistency & Standards
Definition	The application should be consistent and follow design standards for text, audio, navigation, and other elements.
Explanation	Terminology and options should remain the same throughout the application to ensure that users do not get confused about their meaning. If there is a standard way to portray the information or interact with the device, these standards should be used to ensure a seamless experience.
Application feature	User-Cognitive; consistency; learnability; predictability; recognition; User-Usage; availability; context-based; exiting; navigation
Examples	Text should be large enough to read, controls should be based on known interactions (for mobile AR, already established gestures, such as "swipe" to scroll and "pinch" to zoom, should be used), malfunctions (such as lag, drift, and jitter) should be avoided, and menu items should look the same throughout the device/application.
Benefits	A consistent experience will allow the user to learn the device/application quickly and be less confused about their interactions with the device/application.
Problems	Misunderstanding this heuristic could lead to confusion and user errors. When features are inconsistent, users cannot develop clear expectations of how they should interact with the device/application.

Checklist Items

Element	Checklist item	More information and Examples
	Are virtual elements easy to delete or close out of?	AR holograms, 2D images, audio, etc. should be easy for the user to close out of so the application does not become cluttered with irrelevant information.
	Can the user pause the application at any point?	The user should be able to easily stop what they are doing and return to their same spot in the application as they please. If there is an interruption in the real world environment that the user has to attend to, they should **NOT** have to turn off

			and/or restart the application and lose their progress.
		Are all aspects of the device's user interface and/or application (virtual elements, controls, text, etc.) clear and readable?	This may depend on the convergence/accommodation ability of users. Older adult users may have difficulty to see virtual elements based on this and applications should accommodate for this. Microsoft's MR best practices suggests placing holograms 1.25-5m away from the user. Text should be legible enough for users to read. This may be impacted by font size, contrast, typeface, and individual differences (age, vision acuity, color vision deficiencies, etc.). Allow for customizable settings to account for individual differences.
		Are virtual elements sized appropriately?	E.g., Are holograms, 2D images, avatars, etc. large enough that the user is aware of them or small enough that the user does not have to back away to see it. What is appropriate may depend on the application. For

			example, a virtual element may include a full-sized room that the user has to walk around to explore.
		Are virtual elements rendered a reasonable distance away from the user's targeted point?	E.g., the virtual element is placed close enough to the where the user selected that they do not have to move to it, or placed far enough away that the user does not need to back away from it to see or interact with it.
		For mobile devices, are the controls based on known interactions for mobile devices?	E.g., pinch to zoom, swipe to scroll, tap, etc.
		For mobile devices, are landscape **and** portrait mode supported?	The application should be responsive and allow the user to hold the phone either vertically or horizontally while interacting with their application. This allows for flexibility for the user to interact with the application as they like.
		For mobile devices, is the application responsive?	Does the application fit many different types of mobile devices? E.g., tablet, smartphone with "notches" of unusable space on their screen, smartphones that are sized differently, etc.
		Do virtual elements act as the user would	AR content should be consistent with real-

		expect them to in the real world?	world mental models, but do not have to be restrained by them.
		Does the device and/or application avoid lag, delays, jitter, drift, and other forms of virtual element malfunctions?	This refers to malfunctions such as (but not limited to) - time delay between user movement and virtual element movement, virtual elements moving in a jagged or jittery fashion due to frame delay, and virtual elements drifting away from user's input (similar to the drift of cursor when mouse is not being used).
	Navigation	Is the navigation consistent throughout the device and/or application?	Navigation throughout different areas of the device or application should remain consistent regardless of where you begin navigation from. E.g., accessing the area of the application to change the level should remain consistent on level 5 and level 15. And the process in which to access different applications on the device should be consistent whether the user is currently in the settings application or device application store.
		Can the user navigate freely throughout aspects of the device and/or application?	The user should be able to find and use different features of the device and/or

			application as they please.
	Environment	Does the device adjust to the environment it is used in?	E.g., does the screen dim in a darker environment?
		Are environmental requirements clearly defined?	E.g., "use indoors" or "avoid bright lighting"
		Does the device and/or application remind users to be aware of their surroundings?	Often times, AR is created in a way that the user can see the real environment around them at all times. However, users may not always attend to their surroundings, and virtual elements or holograms can occlude potential hazards in the real environment. As such, the application should remind the user to be aware of their surroundings.
	Text	Does the device and/or application avoid jargon?	E.g., specific scientific vocabulary, abbreviations, etc.
		Are sans serif font types used, as appropriate, throughout the device and/or application?	In most instances, fonts such as Arial, Veranda, or Helvetica are sans serif fonts and are easier to read than serif fonts (such as Times New Roman and Halesworth) on screens. However, serif fonts can be easier to read when a large amount of text is presented.

		Is the contrast between the background and text sufficient enough that the text can be read easily under a range of normal lighting conditions?	Text can be difficult to read if there is not enough contrast between it and the background behind it, so it is important to choose a text and background with sufficient contrast. In addition to this, it is difficult to control exactly where an AR device or application will be used. The user could be in a bright environment where text and its background could become washed out, or in a dim environment where the text and background could become too bright.
		If the text background is transparent, is the text visible across different backgrounds and under a range of normal lighting conditions?	It is difficult to control exactly where a user will use an AR device or application. It is possible that they will use it in a room with minimal distractors such as wallpaper, pictures on the wall, machinery, etc. But it is also likely that they will be in a visually busy environment. Text can be difficult to read if there is not enough contrast between it and the background behind it, so it is important to choose a text that is large and clear enough to read in a variety of

			different environmental backgrounds.
		Does the application and/or device use inclusive text and images?	Inclusive text includes using gender-neutral terms, replacing colloquial and culture-specific expressions with plain language, using welcoming language ("you" and "your" rather that "the user"), etc. Images shown should be representative of a diverse population
	Audio	Is the volume adjustable so the user can hear audio, even in noisy environments?	E.g., a user can hear the audio even on a factory floor setting
		Are auditory features understandable?	The user should be able to understand what is being said or depicted through audio output. If spatial audio is being used, the user should be able to identify what direction the audio is coming from.

		Are captions available for auditory features as appropriate?	Include captions for auditory features to increase accessibility. Information such as spoken words, who is speaking, and information that is important for the experience such as sound effects and where the sound came from should be included. Captions should be easy to turn on and customize for readability (e.g., size of text, contrast of text, and font styles). They should also avoid cluttering the screen by showing a limited amount of words on the screen at a time.
Usability attributes	Efficiency; learnability; Memorability; satisfaction		
UX factors	Usable; findable; useful; consistency; workload reduction		
Set of heuristics related	Nielsen's 10; Santos, 2016; Endsley, 2017; Liang, 2018; Gale, 2015; Ko, 2013		

ID	**H6**
Priority	(2) Important
Name	Collaboration
Definition	When sharing an AR/MR space with others, it should be easy to understand what actions are available, what is private vs. public, and communication between users should be seamless.
Explanation	Sometimes AR/MR applications can include multiple users, either with the same abilities or different abilities than each other. It should be clear how communication and teamwork can occur in these spaces (both the abilities and the limitations of available interactions). It is also important to allow users to have a private space in case they do not want to share all content. Finally, avatars should represent users accurately (representative of a diverse population as well as the user's current actions).
Application feature	User-Interaction; satisfaction; efficiency; User-Usage; satisfaction

Examples	Avatars that represent the users, communication includes non-verbal as well as verbal cues to make the experience more natural, the ability to "lock" virtual elements (so users other than the creator of the content cannot delete them).
Benefits	Including avatars that represent a diverse set of users can make the experience enjoyable and more relatable across users. Making interactions and expectations clear between multiple users sharing an experience can increase communication, task success, and enjoyment between users.
Problems	Misunderstanding this heuristic could cause users to become frustrated or mistrustful of a device or application if it is no clear what is shared to others, how to communicate with others, and/or losing autonomy to others' control of the virtual content.
Checklist Items	

Element	Checklist item	More information and Examples
	Are avatars representative of a diverse population?	Virtual representations of real people should be representative of the diverse population that may use them. This could be portrayed through diverse presets or customizability of skin tones, body types, genders, hair texture/color, height, facial shapes, etc.)
	If users are sharing the same virtual space, are virtual landmarks included to help orient users who may be in different physical spaces?	Multi-user communication can be enhanced by having shared landmarks as references. These landmarks may be 3D virtual holograms, 2D elements, etc. For example, users can give instructions to one another by having the same references, "place that block over next to that hologram of a table over there."

				"Do you see the tool? It is next to the axel" etc.
			Is it easy to share virtual content to other users?	Sometimes when collaborating in an application, users want to share additional content to one another (e.g., files, webpages, holograms, video feed, etc.) It should be simple to share this content with multiple users
		Privacy	Does the device and/or application allow for the user to control privacy-related content?	Users' privacy should be protected. This includes their data, personalized settings, application-created content, etc.
			Is sharing virtual content to other users a mutual agreement?	Two-party sharing consent should be encouraged between the sending party and the receiving party. This may be completed with actions such as enabling sharing settings to "on", asking users if they would like to receive virtual content before placing it in their space, having the sending party state that their virtual content is available to others and then letting

			the receiving party download as they please, etc. It is important to let all parties involved know that communication is occurring.
		When collaborating with others, is a private space for the user provided?	It may be necessary for a user to understand what "I" can see vs. "others" can see. Users may not want to share all of their virtual elements with others or may want to view it before sharing with others. A private space can enable this interaction without sharing all content to other collaborators.
		Are avatars and virtual content rendered an adequate distance away from the user to preserve the user's personal space?	Even in virtual environments, where avatars are present, users feel more comfortable having a personal space for themselves to interact with the environment. Avoid avatars and others' virtual objects appearing directly in front of the user, invading this personal space.

		Does the device and/or application allow for the user to easily get to a private "safe place" in the event that other users are making them uncomfortable?	It should be easy and fast for users to get to a safe private space if other users are harassing, bullying, or making them feel uncomfortable. This may be done with a quick key, or shortcut on the device itself or in the application
	Control	Does the application allow users to preserve virtual elements from others users' changes?	E.g., "locking" virtual elements so users other than the creator of the content cannot move them around or delete them. These settings could be global (set for the entire virtual space), for a specific collaborator (e.g., "locking" permissions from one user while allowing another to move objects), or for specific objects.
		Is it clear what another user is referencing using non-verbal cues?	It should be clear from another user's avatar what virtual content they are referencing or interacting with. This could be done via a gesture, head pointing, ray casting, facial expression, head nod, etc.
		Is it clear which virtual elements can be interacted with and which cannot for each user?	It should be clear what can and cannot be interacted with in the moment. Some virtual elements may be "locked" for some users (so they cannot manipulate them). Is

			this clearly communicated to the users through image, color, text, auditory sounds, etc., as appropriate?
	Consistency	Is content consistent across all users, as it is appropriate?	If users are collaborating in the same AR space, they should have the same AR experience across devices and platforms (mobile, desktop, HMD, etc.). However, it may differ based on the purpose of the application (E.g., a professor may have greater access to different controls and content than a student using the same application).
Usability attributes	Satisfaction; efficiency		
UX factors	Usable; communication; accessible		
Set of heuristics related	N/A		

ID	**H7**
Priority	(3) Critical
Name	Comfort
Definition	The application and device should be designed to minimize user discomfort.
Explanatio n	Users should not experience any pain, discomfort, nausea, or disorientation as a result of using AR/MR. The device/application should minimize eye strain, remind the user to take breaks, and provide information about potential risks of discomfort to the user.
Application feature	User-Interaction; direct manipulation; feedback; low physical effort; responsiveness
Examples	Include pop-ups that remind the user to take breaks, design the AR/MR device to be lightweight, allow for the use of protective equipment if it is necessary (such as hard hats, gloves, and/or eye protection), design for diversity (e.g., accommodate for different head sizes, face shapes, hair-

<table>
<tr><td></td><td>styles, corrective lenses, etc.), and avoid unsafe or uncomfortable interactions (e.g., large swiping gesture that involves the movement of the entire arm while in a crowded manufacturing environment, or interactions that force the user to look upward for long periods of time).</td></tr>
<tr><td>Benefits</td><td>Creating a device/application for comfort can include more users to participate in the use of the device/application and make it a more enjoyable experience.</td></tr>
<tr><td>Problems</td><td>Misunderstanding of this heuristic can cause the user to feel uncomfortable and can potentially harm the user.</td></tr>
<tr><td>Checklist Items</td><td>
<table>
<tr><th>Element</th><th>Checklist item</th><th>More information and Examples</th></tr>
<tr><td>Physiological Comfort</td><td>Can the user experience the device and/or application without pain, discomfort, nausea, disorientation, etc. DURING use?</td><td>Users should not be in pain, discomfort, or nauseous as a result of using AR content. Examples of this include eye strain, skin discomfort, pain caused by the heat of the device, muscle strain, headaches, and disorientation. Allow the user to turn off any potentially hazardous functions (e.g., ensuring that flickering images are set to a minimum or can be turned off/reduced for those with epilepsy).</td></tr>
<tr><td></td><td>Can the user experience the device and/or application without pain, discomfort, nausea, disorientation, etc. AFTER use?</td><td>Users should not be in pain, discomfort, or nauseous as a result of using AR content. Examples of this include eye strain, skin discomfort, pain caused by the heat of the device, muscle strain, headaches, and disorientation. Allow the user to turn off any potentially hazardous functions (e.g., ensuring that flickering</td></tr>
</table>
</td></tr>
</table>

				images are set to a minimum or can be turned off/reduced for those with epilepsy).
			Is the device's weight light enough to feel comfortable?	Heavy devices can cause fatigue. This can also differ based on population (e.g., children may not be able to withstand the same weight of a device as an adult) or amount of physical workload as a result of gesture controls or other completing activities when using the device.
			Does the device avoid overheating to the point that it is uncomfortable to use?	Device overheating for a head mounted display or handheld device cause the user to feel uncomfortable using the device and can become dangerous after prolonged use. It is important to avoid overheating the device to the point of discomfort for the user.
			Does the device's accessories and cords avoid hindering work?	The device's cords and accessories should not get in the user's way when completing tasks.
			Are physical interactions with the application safe and comfortable?	It is important to take into account the environment in which the device is being used. A safe gesture in

			a spacious environment may be safe at first, but dangerous when in an industrial setting (e.g., near large manufacturing machines).
		Does the application avoid making the user walk backwards, pull their head back, or push their head downwards to see virtual elements?	If virtual elements appear too close to the user, they may have an instinct to pull their head backwards to "look up", downwards to "look down", or walk backwards. This can cause a hazard as the user could trip over an object that is behind them or strain their neck.
		Do interactions with the device and/or application avoid tiring the user?	User fatigue can cause them to discontinue using the device or applications, and can become dangerous in high risk situations, such as near heavy machinery. Users should not become overly tired or fatigued after using an AR device or application.
	Eye Strain	Does the device and/or application avoid causing the user eye strain?	Eye strain can cause the user discomfort, headaches, and even nausea. It is important to avoid causing the user eye strain in order to provide a safe and enjoyable experience. This can be done by encouraging the user to complete eye calibrations and

			adjusting interpupillary distance (IPD) of the device to match their own.
		Are users reminded to take breaks?	Overextended use can cause eye strain, discomfort, headaches, and even nausea. It is important to remind the user to take breaks to provide a safe and enjoyable experience. This can be done automatically by the application and/or device, or set by user-set options.
	Adaptability	Does the device easily adjust for a diverse set of users?	The device and application should accommodate for as much of a diverse population as possible. This includes device-specific customizability for different head sizes (including inserts or resizing capabilities), face shapes, hair-styles, users with eyeglasses, interpupillary distance (IPD), etc. It also includes customizability of the application such as button remapping, sensitivity of input methods, screen magnification, contrast/luminosity, spatial vs. mono audio, voiceover, captions,

			and disabling non-critical virtual environment content. More examples include allowing users to access native screen readers/voice assistants, and enabling connections with assistive technology devices.
		Does the device accommodate for personal protective equipment?	E.g., It allows for the use of hardhats, gloves, and/or eye protection that comply with safety standards.
Usability attributes	Satisfaction; efficiency		
UX factors	Usable; communication; accessible		
Set of heuristics related	Liang, 2018		

ID	**H8**
Priority	(2) Important
Name	Feedback
Definition	The application and device should provide adequate feedback to the user to explain what is currently going on.
Explanation	When a user or device/application is completing an action, the device/application should provide some sort of feedback (visual, auditory, etc.) to the user that indicates that the action was successful. Likewise, if there is an issue while completing the action, feedback should be given to the user explaining why the action could not be completed.
Application feature	User-Information; defaults; enjoyment; familiarity; hierarchy; multi-modality; visibility; User-Interaction; direct manipulation; feedback; low physical effort; responsiveness
Examples	Include a visual representation while scanning the environment and during loading screens. If an automatic selection occurs (e.g., when a user fails to select an option within a reasonable amount of time) provide the user feedback about the option that was chosen and what to do next.
Benefits	Providing feedback to the user can help the user understand what the device/application is doing and how to proceed. This reduces the amount of

<table>
<tr><td></td><td colspan="3">confusion that they may have when interacting with the device/application and sets expectations about how it should perform.</td></tr>
<tr><td>Problems</td><td colspan="3">Misunderstanding of this heuristic can cause confusion and frustration from the user. It may also result in errors that users may not understand how to recover from.</td></tr>
<tr><td>Checklist Items</td><td>Element</td><td>Checklist item</td><td>More information and Examples</td></tr>
<tr><td></td><td></td><td>Does the device and/or application provide feedback on its status?</td><td>E.g., loading screen, representation when scanning the environment, etc.</td></tr>
<tr><td></td><td></td><td>Does the device and its accessories provide feedback about battery levels and charging state?</td><td>Users should be able to easily see the status of their device and accessories to be able to estimate how long they can use them before recharging.</td></tr>
<tr><td></td><td></td><td>Is it clear how user data is collected, stored, used, and protected?</td><td>Users should be able to recognize what information they give is kept private and what is not. This is sometimes found within help sections, additional documentation for the application, privacy settings, or within user prompts upon first use (e.g., "This app would like to access the camera to enable AR"). This should be disclosed in plain language and ask for the user's informed consent to collect data. Applications and devices should protect user data wherever possible (encoding data, avoid capturing unnecessary data, and using additional</td></tr>
</table>

				techniques such as user authentication and following laws such as HIPPA laws as appropriate).
		Feedback after input or selection	Does the device and/or application provide feedback for user input?	When the user completes an input, the device should give the user feedback (e.g., auditory "click", tactile vibration on controller, color change on the screen, etc.)
			Does the device and/or application respond quickly to user input?	E.g., taps, voice input, movement, etc.
			Does the device and/or application provide the user feedback after automatic selections?	Automatic selections can occur when a user fails to provide input after a point of time. (E.g., a user fails to select either option a or b within 60 seconds, so the application automatically chooses option a for them and continues on to the next section)
			If an automatic selection occurs, does the device and/or application suggest what to do next?	Automatic selections can occur when a user fails to provide input after a point of time. (E.g., a user fails to select either option a or b within 60 seconds, so the application

			automatically chooses option a for them and continues on to the next section)
Usability attributes	Satisfaction; desirable; efficiency		
UX factors	Simplicity; efficiency; usable; communication; accessible		
Set of heuristics related	Nielsen's 10; Santos, 2016; Liang, 2018; de Paiva Guimarães & Martins, 2014		

ID	**H9**		
Priority	(3) Critical		
Name	User Interaction		
Definition	All interactions that the user has with the device/application should be simple, easy to understand, and easy to complete.		
Explanation	User interactions should be easy to learn, and quick, safe, and comfortable to perform. Different ways of interaction should be available so the user can choose the method that works best for them. The user should feel confident about their control of the application and the virtual elements within it.		
Application feature	User-Interaction; direct manipulation; feedback; low physical effort; responsiveness; User-Support; error management; help & documentation; personalization; user control		
Examples	Allow for multiple ways to interact with an object (e.g., allow the user to "resize" an object through a voice command, one-handed pinch motion, two-handed gesture, or menu option).		
Benefits	Quick and easy to learn user interactions can decrease the amount of cognitive workload and number of errors that a user experiences while also increasing efficiency and user satisfaction.		
Problems	Misunderstanding of this heuristic can cause the device/application to become more difficult to learn and frustrating for the user. It also may result in the abandonment of use.		
Checklist Items	**Element**	**Checklist item**	**More information and Examples**
	User Control	Does the user feel in control?	The user should feel like they have a direct influence on the device and/or application and that they can control how it responds to their input.

		Are user interactions simple and easy to understand?	User interactions should be simple and easy so they do not confuse the user, or cause the user to repeatedly check tutorials/instructions to remember them.
		Does the device and/or application include multiple forms of interaction so users can choose based on ability, preference, & skill?	Allowing the user to interact how it works best for them makes a more seamless interaction experience, this can be done by implementing multiple forms of input (e.g., mouse & keyboard input, taps, gesture controls, voice commands, etc.) and/or different methods of using these input types (e.g., stationary such as sitting or standing, dynamic while moving, close or far away, ambidextrous gesture input, customization of sensitivity of input for those who may have difficulty with fine motor control, providing compatibility with assistive technologies, etc.).

		Are the forms of interaction direct when it is appropriate to use this form of interaction?	Direct interaction can be easier to use and understand than indirect interaction (e.g., being able to manipulate a 3-D object by touching it instead of through a menu or voice command next to the object). However, direct interaction cannot always be used (e.g., if an employee needs both hands to be able to interact with an engine part, but also needs to be able to look at an AR application. In this case, voice (indirect interaction) would be a better form of interaction rather than having to move their hands between the engine and AR device (direct interaction))
		Does the device and/or application avoid interactions that force the user to make large or sudden movements?	Large or sudden movements can be hazardous. The user could pull a muscle or strike an object in a crowded environment, harming themselves or others around them.
		Does the device and/or application accommodate for the user to complete other necessary real-world tasks?	Keep in mind that this may be one-handed, hands-free interaction, or N/A based on what types of tasks the user is required/not required to complete alongside the AR experience.

	Device and/or Application Performance	Does object manipulation work well in all instances?	If there are multiple ways to interact with an object (e.g., "select" an object through a voice command, one-handed pinch motion, two-handed gesture, gaze, or menu option), ensure that these interactions work well in all instances (e.g., if user is walking, interacting with another object, holding a mobile device with one hand, etc.).
		Do virtual elements adapt to the users' position appropriately?	The device and/or application should be able to adapt virtual elements in a useful manner. If the virtual content needs to be world-locked, allowing the user to walk around it, the device and/or application should support this. This also applies to content that should be head-locked (follows the user's head position) or blended between world-locked and head-locked. The device and/or application should support interaction methods that allow the user to make sense of the content.
		Does the device and/or application avoid input overloading by assigning distinct	It is best to avoid many functions for one button, gesture, or other input method as this can become

		functions to buttons or gestures?	confusing and difficult for users to remember.

Usability attributes	Effectiveness; satisfaction; errors; efficiency
UX factors	Usefulness; error prevention & handling; usable; communication; accessible
Set of heuristics related	Endsley, 2017; Nielsen's 10

ID	**H10**
Priority	(2) Important
Name	Recognition Rather than Recall
Definition	The application should be designed in a way that promotes the use of recognition rather than recall to minimize the user's memory load.
Explanation	Providing too much information or novel concepts that need to be remembered to either use the device/application or complete a later task is cognitively taxing for the user. The user should not have to remember information from one part of the device/application to another. Encouraging the use of recognition by providing guidance and recognizable elements rather than forcing the user to recall all previous information will ease cognitive load.
Application feature	User-Cognitive; consistency; learnability; predictability; recognition
Examples	If virtual elements are off-screen, include arrows (or other form of identification) to grab the user's attention to it. Virtual tags (such as those that label machine parts or other physical or virtual objects) should accurately follow the parts as the user moves and manipulates them.
Benefits	Minimizing workload can reduce user's mental fatigue, help users complete tasks effectively and efficiently, and make it easier for users to learn the device/application quickly.
Problems	Misunderstanding of this heuristic can lead to higher cognitive workload and can result in user errors and frustration.
Checklist Items	

Element	Checklist item	More information and Examples
Virtual elements	Are virtual elements and icons self-explanatory (does their form communicate function)?	E.g., Instead of a button that states the word "save", an icon of a floppy disk can be used.
	Are virtual elements and controls placed	E.g., the word "rotate" near a ball will rotate

		near objects they reference?	the ball, not another AR element off screen.
		If a virtual element is related to an object that is in motion, is the virtual element tightly coupled with object in motion appropriately?	E.g., Virtual tags that label real machine parts should accurately follow the parts as the user moves and manipulates them.
		Are virtual elements that are outside of the field of view easy to find?	E.g., arrows point in the direction of a virtual AR object off screen
	User	Are available user actions identifiable?	Users should be able to recognize and identify all of the actions available to them in a device's user interface and/or application.
		If voice commands are included, are text labels for voice commands given?	Text labels can help clarify what the user should say for specific voice commands, and allow the user to get a quick glance of possible commands for the device or application.
Usability attributes	Efficiency; learnability; memorability		
UX factors	Usable; findable; useful; consistency; workload reduction		
Set of heuristics related	Nielsen's 10; Kalalahati, 2015; Endsley, 2017; Liang, 2018; Ganapathy, 2013		

ID	**H11**
Priority	(1) Useful
Name	Device Maintainability
Definition	The device should be designed in a way that makes it easy to maintain. This includes reusability, storage, cleaning, and the ability to fix/replace parts.
Explanation	The device should be able to be used multiple times. During this reuse, it is possible that the device may get dirty and parts may break. It is important that the process to clean and fix/replace parts is easy (and occurs infrequently) so the user can trust the reliability of the device.

<table>
<tr><td>Application feature</td><td>User-Usage; availability; context-based; exiting; navigation</td></tr>
<tr><td>Examples</td><td>If a part (such as a strap) breaks, it should be easy to purchase and install a new one.</td></tr>
<tr><td>Benefits</td><td>It is very likely that parts will need to be cleaned or replaced after many instances of reuse. User satisfaction can increase if you give control to users by giving them the knowledge and access to clean and replace parts themselves.</td></tr>
<tr><td>Problems</td><td>Misunderstanding of this heuristic can cause potential damage to the device and user frustration.</td></tr>
<tr><td>Checklist Items</td><td>
<table>
<tr><th>Element</th><th>Checklist item</th><th>More information and Examples</th></tr>
<tr><td>Device protection</td><td>Is the device sturdy enough to withstand multiple uses?</td><td>The device should be sturdy enough to avoid damage from normal use.</td></tr>
<tr><td></td><td>Does the device have a sturdy storage case?</td><td>The device should be stored securely to prevent any damage.</td></tr>
<tr><td>Device upkeep</td><td>Is it easy to clean the lenses, cameras, and other components on the device?</td><td>It is important to be able to clean and sanitize the device, especially if multiple people will use the device.</td></tr>
<tr><td></td><td>Are device parts fixable and replaceable as needed?</td><td>E.g., if a strap breaks, it should be easy to replace and install a new one</td></tr>
<tr><td></td><td>Does the device's battery life last long enough to perform necessary tasks of the application?</td><td>The user should not have to interrupt their task to change or recharge the device's battery. If the necessary tasks take two hours to complete, the device's battery should also last for two hours.</td></tr>
</table>
</td></tr>
<tr><td>Usability attributes</td><td>Efficiency; satisfaction</td></tr>
<tr><td>UX factors</td><td>Usable</td></tr>
<tr><td>Set of heuristics related</td><td>N/A</td></tr>
</table>

Appendix F

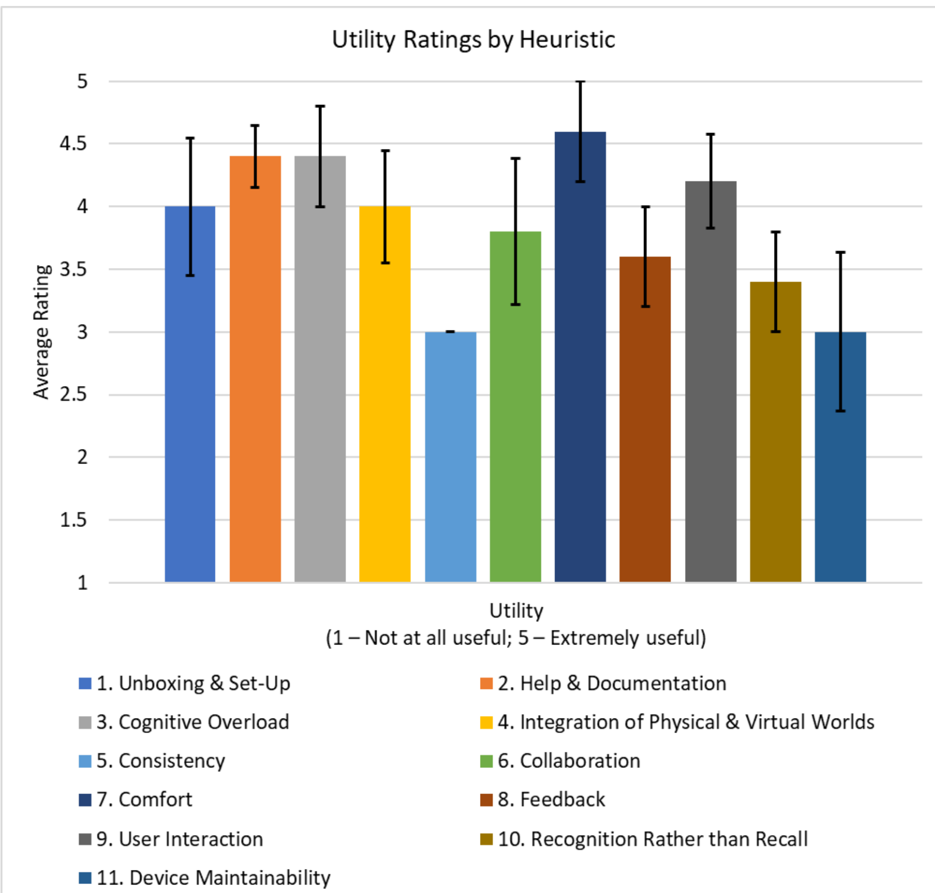
Average Ratings of Each Heuristic Given by Expert Evaluators

Utility

Question given to expert evaluators: “How useful is the heuristic?” (1 – Not at all useful; 5 – Extremely useful).

Figure F 1

Utility Ratings by Heuristic as Rated by Expert Evaluators



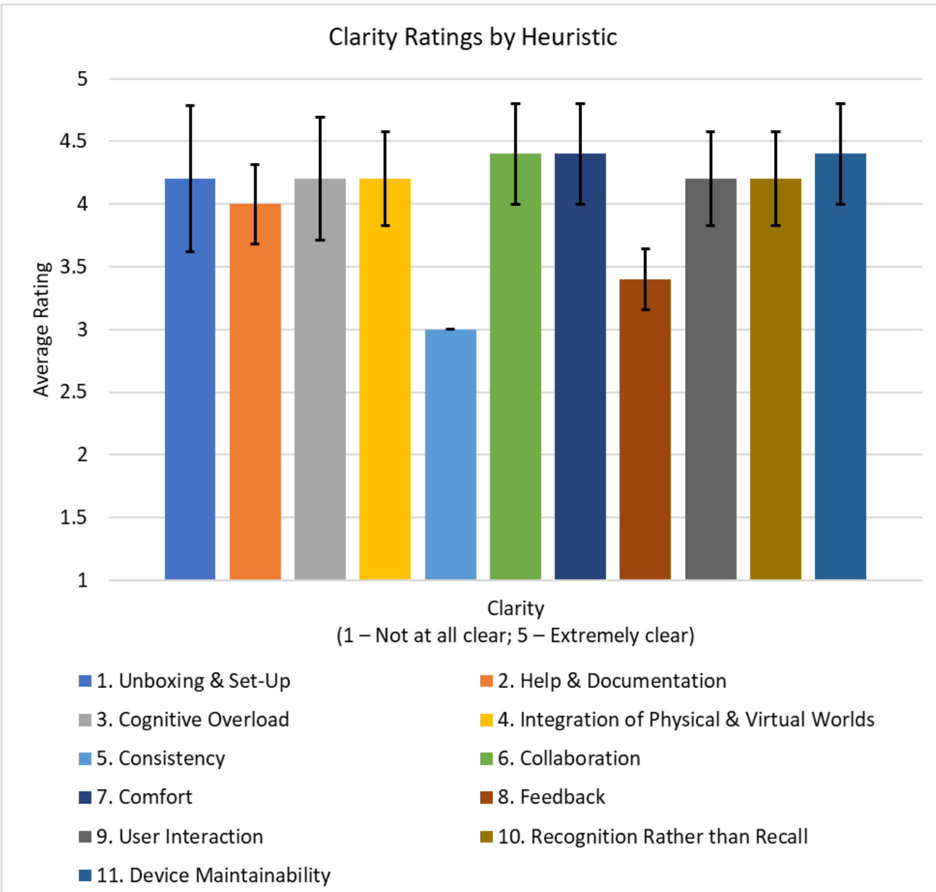
Note: Error bars represent the standard error (SE).

Clarity

Question given to expert evaluators: "How clear is the heuristic?" (1 – Not at all clear; 5 – Extremely clear).

Figure F 2

Clarity Ratings by Heuristic as Rated by Expert Evaluators



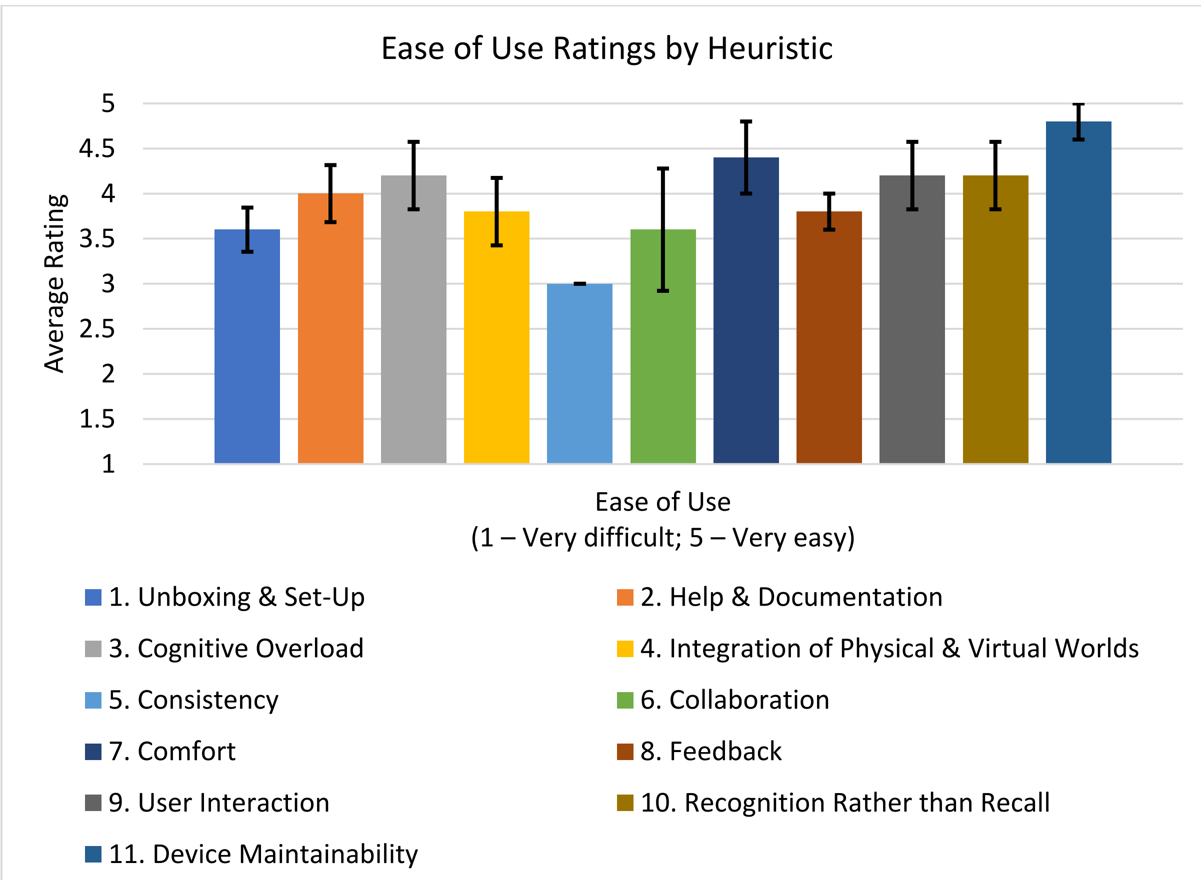
Note: Error bars represent the standard error (SE).

Ease of Use

Question given to expert evaluators: "How easy it was to associate identified problems with the heuristic?" (1 – Very easy; 5 – Very difficult).

Figure F 3

Ease of Use Ratings by Heuristic as Rated by Expert Evaluators



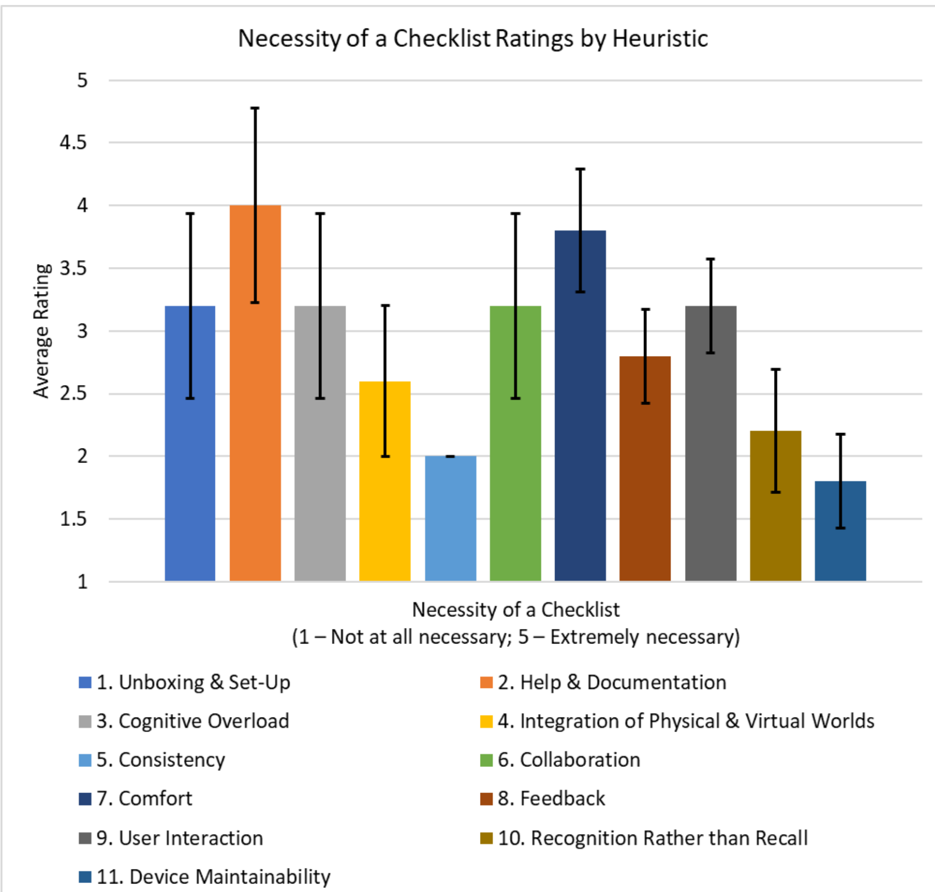
Note: Error bars represent the standard error (SE).

Necessity of a Checklist

Question given to expert evaluators: "How necessary is it to complement the heuristic with a checklist?" (1 – Not at all necessary; 5 – Extremely necessary).

Figure F 4

Necessity of a Checklist Ratings by Heuristic as Rated by Expert Evaluators



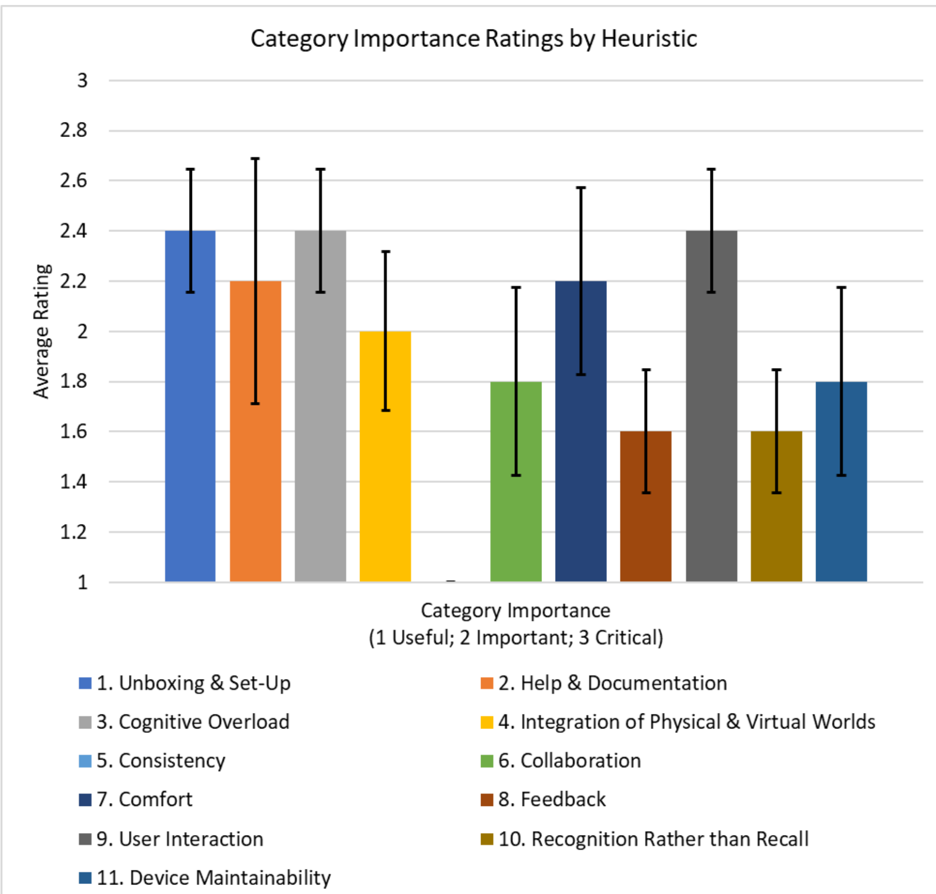
Note: Error bars represent the standard error (SE).

Category Importance

Question given to expert evaluators: "How important is the heuristic?" (1) = Useful: Heuristic further improves the usability/UX; (2) = Important: Heuristic evaluates a relevant aspect; (3) = Critical: Heuristic evaluates a crucial aspect.

Figure F 5

Category Importance Ratings by Heuristic as Rated by Expert Evaluators



Note: Error bars represent the standard error (SE).

Appendix G

Detailed Descriptions of Applications and Devices

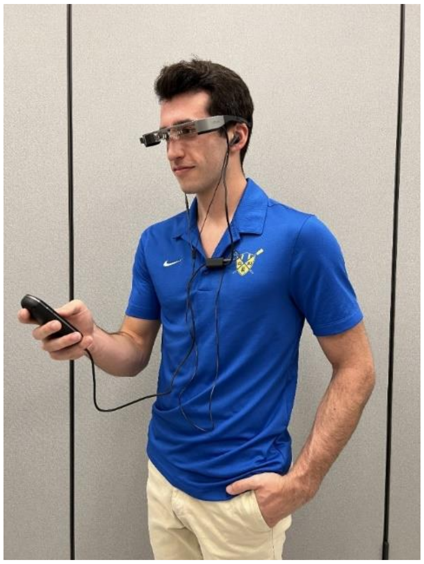
Epson Moverio BT-300 & a Variety of Applications

Heuristic evaluators and user testing participants were asked to use the Epson Moverio BT-300, an AR smart glasses device. This is a tethered lightweight device that uses optical see-through technology to show AR content on the user's screen. It includes a headset and a controller that are connected by a cord. Evaluators and participants interacted with this device by using the controller, and through head-tracking. Its technical specifications were described above in Table 4. An image of the device is shown in Figure G1.

A variety of apps were used on this device. This included: Asteroid Fighter (a game that required players to fly towards and shoot to destroy asteroids using their head movements and the controller), Age of Diamonds (a brick-breaker style of game that required players to control a ball and break blocks on the screen using their head movements), Camera (a native camera application on the device; photos were taken through controller input), Gallery (a native photo gallery on the device; heuristic evaluators and participants accessed this through controller input), Browser (a native internet browser application on the device; evaluators and participants were asked to watch a video and accessed this through controller input).

Figure G 1

A User Wearing the Epson Moverio BT-300 Smart Glasses



Procedure

Heuristic evaluators and user testing participants were asked to come to a research lab on campus, and upon arrival participants were given an informed consent form. Participants signed the consent form before participating. First, participants were asked to fill out a demographic questionnaire. Next, the facilitator gave evaluators and participants an introduction of the device that covered: parts of the device, the controls of the device, providing an instruction manual, and informing evaluators and participants that they could take breaks as needed. Participants were also notified that their screen was being shared to a facilitator's computer screen so the facilitator could help troubleshoot as needed.

Evaluators and participants were asked to wear the device for a few minutes and practice the controls. They were then asked to complete a series of tasks:

1. Play a game called "Asteroid Fighter" for 5 minutes

2. Play the first 5 levels of a game called “Age of Diamonds”
3. Take a photo using the device
4. Review the photo that they just took
5. Look up a movie trailer for the movie “Elementals” on the internet browser and watch the trailer.

Heuristic evaluators were given the opportunity to complete the evaluation after using the application. They were also told that they could take notes while using the application, but to mostly focus on completing the tasks and experiencing the application.

User testing participants were asked to give their first impressions of the application and to rate how difficult they expected the application will be to use (using the task difficulty scale). During the tasks, they were asked to speak aloud if they had comments about the application. After each task was completed, participants were asked to rate how difficult it was to complete the task and if they had any comments about the task. After all tasks were completed, participants complete a series of questionnaires on a computer (SUS, NASA-TLX Raw, SSQ, and Eyestrain Questionnaire), rated how difficult it was to use the application overall (using the task difficulty scale), and provided any comments they had about using the device and application. Participants were given a debrief upon completion of the study and given compensation. The study took approximately 60 minutes to complete.

Magic Leap 1 & Wayfair Spaces

Heuristic evaluators and user testing participants were also asked to use the Magic Leap 1, a MR HMD. This is a tethered device that uses optical see-through technology to show MR content on the user's screen. It includes a headset and a power pack that are connected by a cord, and a controller. Evaluators and participants interacted with the device and application by using the Magic Leap 1 controller, though this device does also have the capacity to be controlled through eye-tracking, hand gesture control, and voice. Its technical specifications were described above in Table 4.

Heuristic evaluators and user testing participants completed tasks using the Wayfair Spaces application. Wayfair Spaces is a free interior design and planning app. Using this app, users are able to virtually try out available Wayfair products (e.g., rugs, sofas, chairs, artwork, and décor) in their real space before deciding to purchase the Wayfair product. This application was initially released in 2018. A screenshot of this application is shown in Figure G2.

Figure G 2

Wayfair Spaces Application

Note: Image from Magic Leap (n.d.b.). *Wayfair Spaces.* https://world.magicleap.com/en-us/details/com.wayfair.spaces

Procedure

Heuristic evaluators and user testing participants were asked to come to a research lab on campus, and upon arrival participants were given an informed consent form. Participants signed the consent form before participating. First, participants were asked to fill out a demographic questionnaire. Next, the facilitator gave evaluators and participants an introduction of the device that covered: parts of the device, the controls of the device, providing an instruction manual, and informing evaluators and participants that they could take breaks as needed. Participants were also notified that their screen was being shared to a facilitator's computer screen so the facilitator could help troubleshoot as needed.

Heuristic evaluators and participants were asked to wear the device for a few minutes and practice the controls. They completed a step-by-step fitting guide and visual calibration. They were then asked to open the Wayfair application and were asked to complete a series of tasks with the Wayfair Spaces application:

1. Complete the Wayfair Spaces tutorial
2. Choose a space to pick virtual objects from
3. Place at least 4 virtual objects into the physical room (one chair, one table or shelf, one object that will sit on that table or shelf, and one wall item)
4. Copy and place a virtual object
5. Move and rotate a virtual object
6. Find the name of a virtual object
7. Swap a virtual object to a different style
8. Place another virtual object from a different scene

9. Place another virtual object from the items tab.

Heuristic evaluators were given the opportunity to complete the evaluation after using the application. They were also told that they could take notes while using the application, but to mostly focus on completing the tasks and experiencing the application.

User testing participants were asked to give their first impressions of the application and to rate how difficult they expected the application will be to use (using the task difficulty scale). During the tasks, they were asked to speak aloud if they had comments about the application. After each task was completed, participants were asked to rate how difficult it was to complete the task and if they had any comments about the task. After all tasks were completed, participants complete a series of questionnaires on a computer (SUS, NASA-TLX Raw, SSQ, and Eyestrain Questionnaire), rated how difficult it was to use the application overall (using the task difficulty scale), and provided any comments they had about using the device and application. Participants were given a debrief upon completion of the study and given compensation. The study took approximately 60 minutes to complete.

Magic Leap 2 & Fan Blade Replacement Application

Heuristic evaluators and user testing participants were asked to use the Magic Leap 2, a MR HMD. This is a tethered device that uses optical see-through technology to show MR content on the user's screen. It includes a headset and a power pack that are connected by a cord, and a controller. Participants interacted with the device and application by using the Magic Leap 1 controller, hand gesture control, and voice. This device does also have the capacity to be controlled through eye-tracking. Its technical specifications were described above in Table 4.

Evaluators and participants completed tasks with the Fan Blade Replacement application. The Fan Blade Replacement application was developed by the Extended Reality (XR) lab Embry-Riddle Aeronautical University. This application uses MR to step aircraft maintenance students through the process of removing and reinstalling General Electric TF34 fan blades. Currently, aircraft maintenance students are required to follow a procedure that is written out on a paper manual. No images are included in this manual. The Fan Blade Replacement application presents the same text as the manual, but also superimposes 3D virtual content onto the TF34 GE-400B aircraft engine. As the application explained to the user how to complete each step, the user was required to physically complete the step on the aircraft engine.

A TF34 GE-400B turbofan aircraft engine was used to overlay virtual content on top of. This type of engine is commonly used on the A-10 Thunderbolt II, S-3 Viking, and RQ-170 Sentinel aircraft. The specific engine that was used for this study was located in an aviation maintenance lab on campus and was secured using an engine test stand. Images of this application and a user using the application are shown in Figures G3 and G4.

Figure G 3

Fan Blade Replacement Application

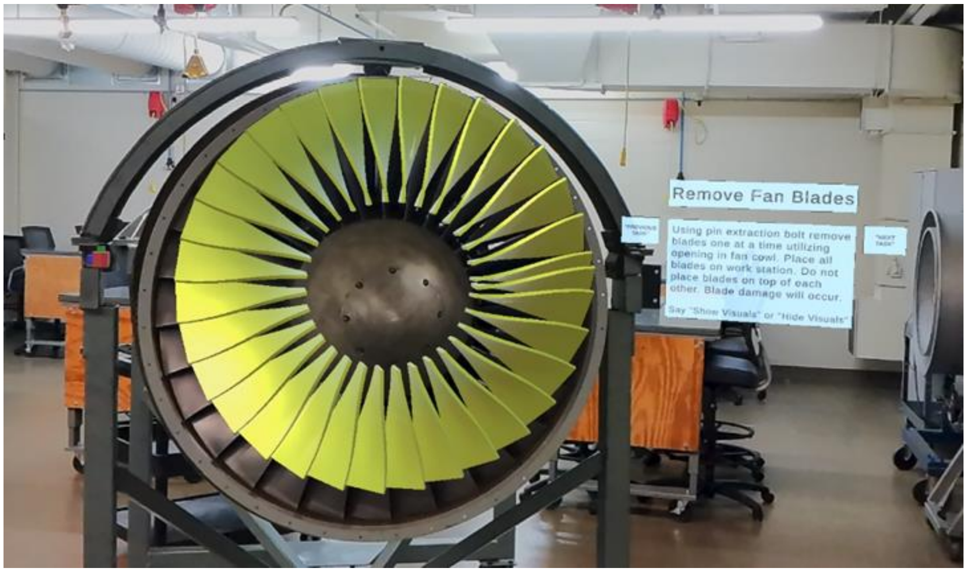


Figure G 4

A User Wearing the Magic Leap 2 While Using the Fan Blade Replacement Application

Procedure

Heuristic evaluators and user testing participants were asked to come to an aviation maintenance lab on campus, and upon arrival participants were given an informed consent form. Participants signed the consent form before participating. First, participants were asked to fill out a demographic questionnaire. Next, the facilitator gave evaluators and participants an introduction of the device that covered: parts of the device, the controls of the device, and informing evaluators and participants that they could take breaks as needed.

Evaluators and participants were asked to wear the device for a few minutes and practice the controls. Then, they completed a step-by-step fitting guide and visual calibration. After, they were asked to open the Fan Blade Replacement application and to complete a series of tasks:

1. Loosen the bolts on the front cover of the nose cone
2. Remove the front cover of the nose cone
3. Loosen the bolts on the rear cover
4. Remove the rear cover
5. Remove the fan blade cover
6. Remove a fan blade
7. Identify the position of the fan blade
8. Reinsert the fan blade (evaluators and participants were only required to remove and replace one fan blade)
9. Place and secure the fan blade cover

10. Place and secure the rear cover
11. Place and secure the front cover

Heuristic evaluators were given the opportunity to complete the evaluation after using the application. They were also told that they could take notes while using the application, but to mostly focus on completing the tasks and experiencing the application.

User testing participants were asked to give their first impressions of the application and to rate how difficult they expected the application will be to use (using the task difficulty scale). During the tasks, they were asked to speak aloud if they had comments about the application. After each task was completed, participants were asked to rate how difficult it was to complete the task and if they had any comments about the task. After all tasks were completed, participants complete a series of questionnaires on a computer (SUS, NASA-TLX Raw, SSQ, and Eyestrain Questionnaire), rated how difficult it was to use the application overall (using the task difficulty scale), and provided any comments they had about using the device and application. Participants were given a debrief upon completion of the study and given compensation. The study took approximately 120 minutes to complete.

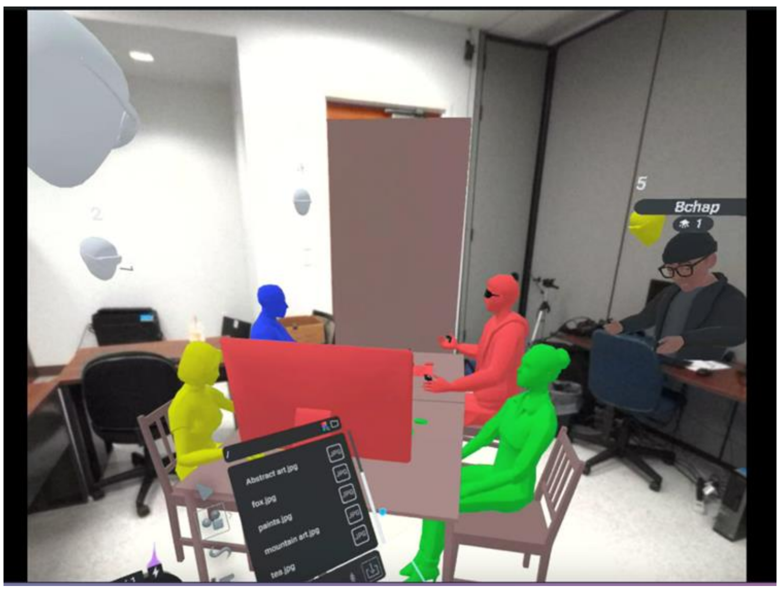
Meta Quest Pro & Shapes XR

The Meta Quest Pro is an XR HMD. This is a tethered device that uses video see-through technology to show AR and MR content on the user's screen. It can also show VR content on the user's screen. It includes a headset and two controllers. Heuristic evaluators and user testing participants interacted with the device and application by using the Meta Quest Pro controllers, though this device does also have the capacity to be controlled through eye-tracking, hand gesture control, and voice. Its technical specifications were described above in Table 4.

ShapesXR is a free design and collaboration app. Using this app, users are able to create 3D prototypes in their real space. This application was initially released in 2021. An image of this application is shown in Figure G5.

Figure G 5

ShapesXR Application



Procedure

Heuristic evaluators and participants were asked to meet the research facilitator in a lab on campus and upon arrival, participants were given an informed consent form. Participants signed the consent form before participating. Heuristic evaluators and user testing participants were asked to wear the device for a few minutes and practice the controls. They completed a step-by-step fitting guide and visual calibration. They were then asked to open the ShapesXR application. Before beginning tasks with the application, they watched an explanatory video about the application. They were then asked to complete a series of tasks with the ShapesXR app:

1. Complete a series of tutorials for the application
2. Work together with another user (this was a confederate researcher located in another room on campus) to create a prototype of a shared office space that satisfied the following requirements:
    a. 4 desks
    b. 4 chairs
    c. 4 computers with an image on each of the screens
    d. Fun decorations
    e. They were also given information about the fake employees they were prototyping the space for to give them a starting point:
        i. Michael loves video games and the color red.
        ii. Emma enjoys cars and the color yellow.
        iii. Tim is an artist whose favorite color is blue.
        iv. Ava likes her morning cup of tea and the color green.

A confederate researcher played the role of another evaluator and worked together with the evaluator or participant to complete task 2. They made sure that the heuristic evaluator encountered the following:

a. Grab a virtual object from another person.

b. Hand a virtual object to another person.

c. Place a virtual object.

d. Recolor a virtual object.

e. Make a virtual object bigger or smaller.

f. Place an image.

g. Delete a virtual object or image.

h. Take photos of the virtual space after completing their prototype.

i. Have an equal amount of input when deciding how to create the prototype and when actively creating the prototype.

j. Experience another user changing a virtual object without permission or discussing beforehand.

Heuristic evaluators were given the opportunity to complete the evaluation after using the application. They were also told that they could take notes while using the application, but to mostly focus on working together with their partner to complete the task and experience the application.

User testing participants were asked to give their first impressions of the application and to rate how difficult they expected the application will be to use (using the task difficulty scale). During the tasks, they were asked to speak aloud if they had comments about the application, but to mostly focus on working together with their partner to complete the task and experience the application without interruption. After all tasks were completed, participants were asked to complete a series of questionnaires on a computer (SUS, NASA-TLX Raw, SSQ, and Eyestrain Questionnaire), rated how difficult it was to complete each of the tasks and experiences shown in 1-2 and a-j above, asked how difficult it was to use the application overall (using the task difficulty scale), and asked to provide any comments they had about using the device and application. Participants were given a debrief upon completion of the study and given compensation. The study took approximately 120 minutes to complete.

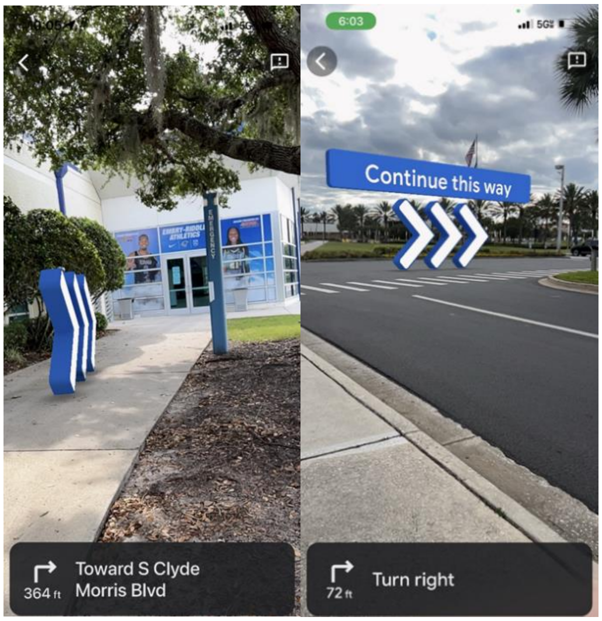
Mobile Phone & Google Maps Live View

Heuristic evaluators and user testing participants were asked to use mobile phones for the Google Maps Live View application. Either an iPhone 13 Pro or an Android Galaxy S8 were given to evaluators and participants depending on if they typically used an iOS or Android phone. Evaluators and participants interacted with the device and application by tapping on the screen and walking in the real environment. Its technical specifications were described above in Table 4.

Evaluators and participants completed tasks using the Google Maps Live View application. Google Maps Live View is a feature in the Google Maps app. When navigating in Google Maps, users can turn on this feature and see AR directions in their real space. This feature was initially released in 2019. An example of what a user may see when using this application is shown in Figure G6.

Figure G 6

Google Maps Live View Application



Procedure

Heuristic evaluators and participants were asked to meet the research facilitator outside of a building on campus (the ICI Center) and upon arrival, participants were given an informed consent form. Participants signed the consent form before participating. First, participants were asked to fill out a demographic questionnaire. Next, the facilitator gave an introduction of the device that covered: parts of the device, the controls of the device, and informing participants that they could take breaks as needed. The facilitator walked alongside the evaluator and participant as they completed tasks. Either an iPhone 13 Pro or an Android Galaxy S8 were given to evaluators or participants depending on if they typically used iOS or Android for their daily phone. Three user testing participants were given an Android phone to use and six were given an iPhone. Three heuristic evaluators were given an Android phone to use and three were given an iPhone.

Heuristic evaluators and participants were asked to open the Google Maps application and enable Live View. They were then asked to complete a series of tasks with the app:

1. Start the AR walking directions to the Eagle Fitness Center across campus
2. Follow the AR walking directions until we reach the crosswalk
3. Cross the street with the AR walking directions
4. Arrive to the final location and end the walking directions

This journey was 0.3 miles. Heuristic evaluators were given the opportunity to complete the evaluation after using the application. They were also told that they could take notes while using the application, but to mostly focus on completing the tasks and experiencing the application.

User testing participants were asked to give their first impressions of the application and to rate how difficult they expected the application will be to use (using the task difficulty scale) before completing tasks. During the tasks, they were asked to speak aloud if they had comments about the application. After each task was completed, participants were asked to rate how difficult it was to complete the task and if they had any comments about the task. After all tasks were completed, participants complete a series of questionnaires on a computer (SUS, NASA-TLX Raw, SSQ, and Eyestrain Questionnaire), rated how difficult it was to use the application overall (using the task difficulty scale), and provided any comments they had about using the device and application. Participants were given a debrief upon completion of the study and given compensation. The study took approximately 60 minutes to complete.

Appendix H

Usability Study Task Success, Difficulty, and Questionnaire Results

Epson Moverio BT-300 & a Variety of Applications

Task Success and Difficulty

All eight participants successfully completed four out of the five tasks. The one task that participants struggled the most with was playing the Asteroid Fighter game. This task success rate is shown in Figure H1. This was also rated the most difficult task by participants (M = 2.5, SD = 1.2). This was mostly due to this application not having instructions to explain to the user how to play the game, and users continuously pressed the wrong buttons on the controller, exiting the application during the game. Difficulty for each task is shown in figure H2. Overall, the device was about as difficult to use (M = 3.5, SD = 1.07) as participants expected it to be at the beginning of the study (M = 3.38, SD = 0.74), this is shown in Figure H3.

Figure H 1

Epson Moverio BT-300: Task Success Rate

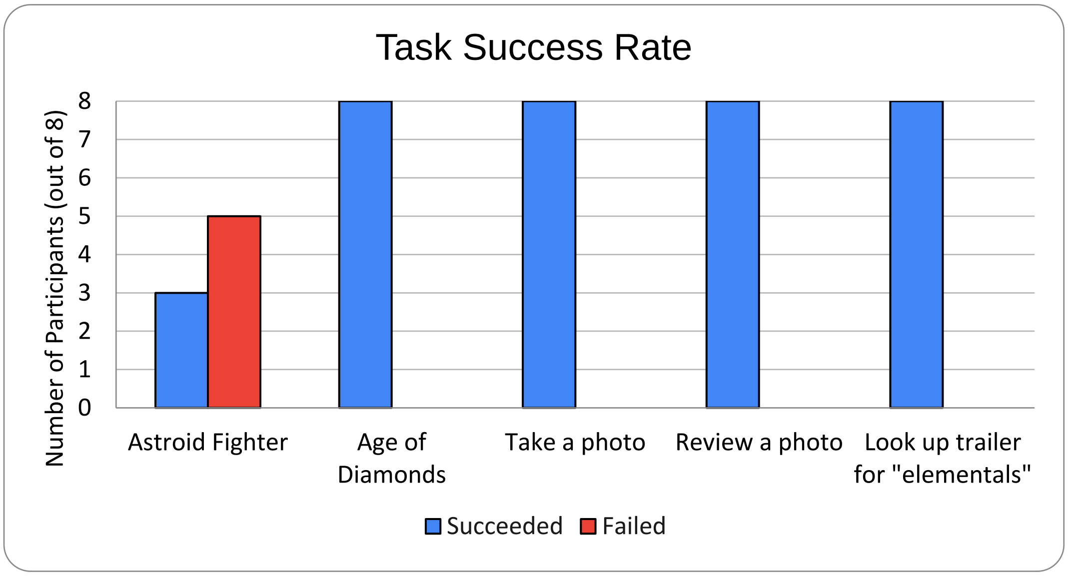
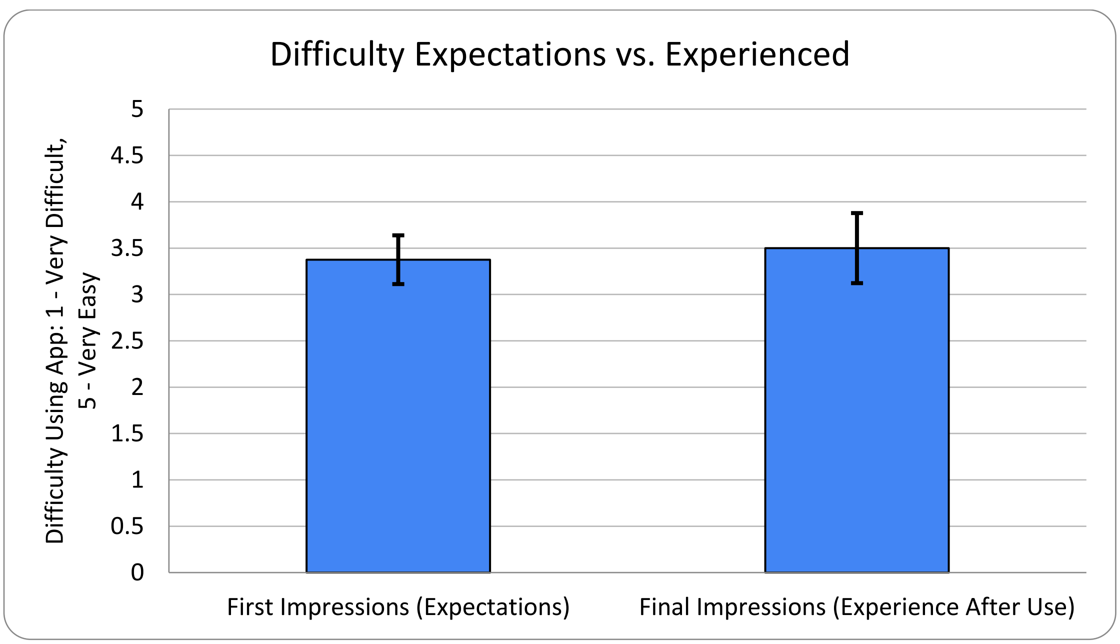


Figure H 2

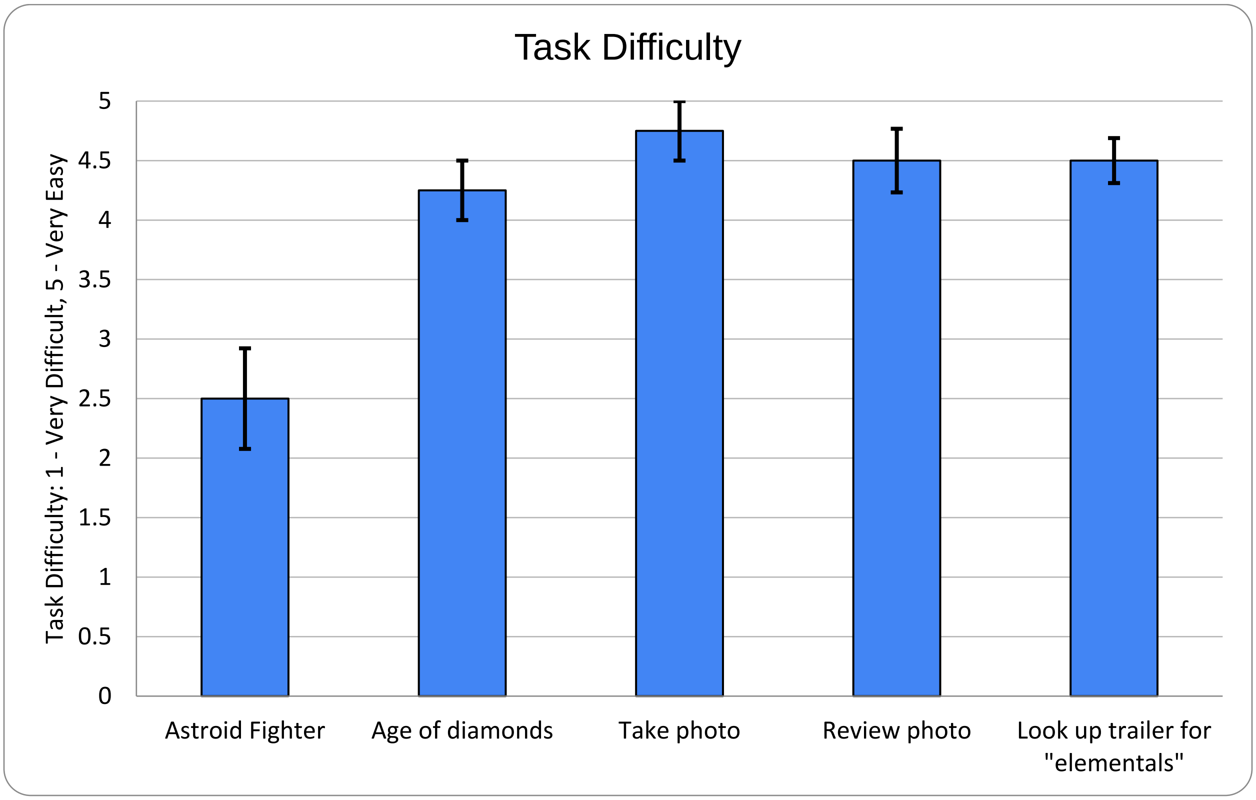
Epson Moverio BT-300: Device Difficulty Ratings Expectations vs. Actual Experience



Note: Error bars represent the standard error (SE).

Figure H 3

Epson Moverio BT-300: Difficulty Ratings by Task



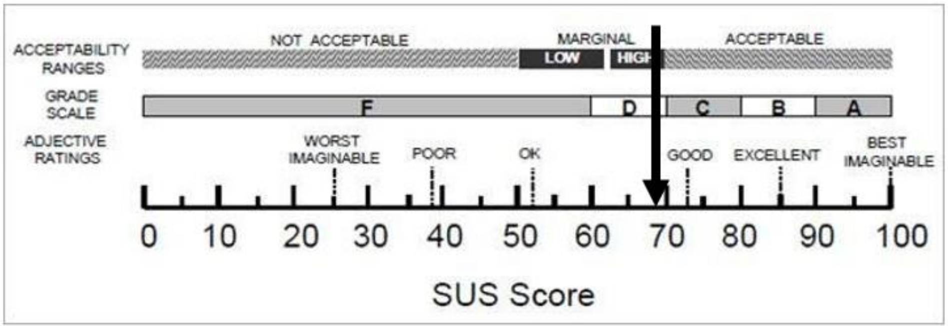
Note: Error bars represent the standard error (SE).

System Usability Scale (SUS)

SUS scores averaged to a "Good" rating *(M* = 69.06, *SD* = 21.59), meaning that overall, the device and application was a good experience for users to use. This is shown in Figure H4.

Figure H 4

Epson Moverio BT-300: System Usability Scale Scores



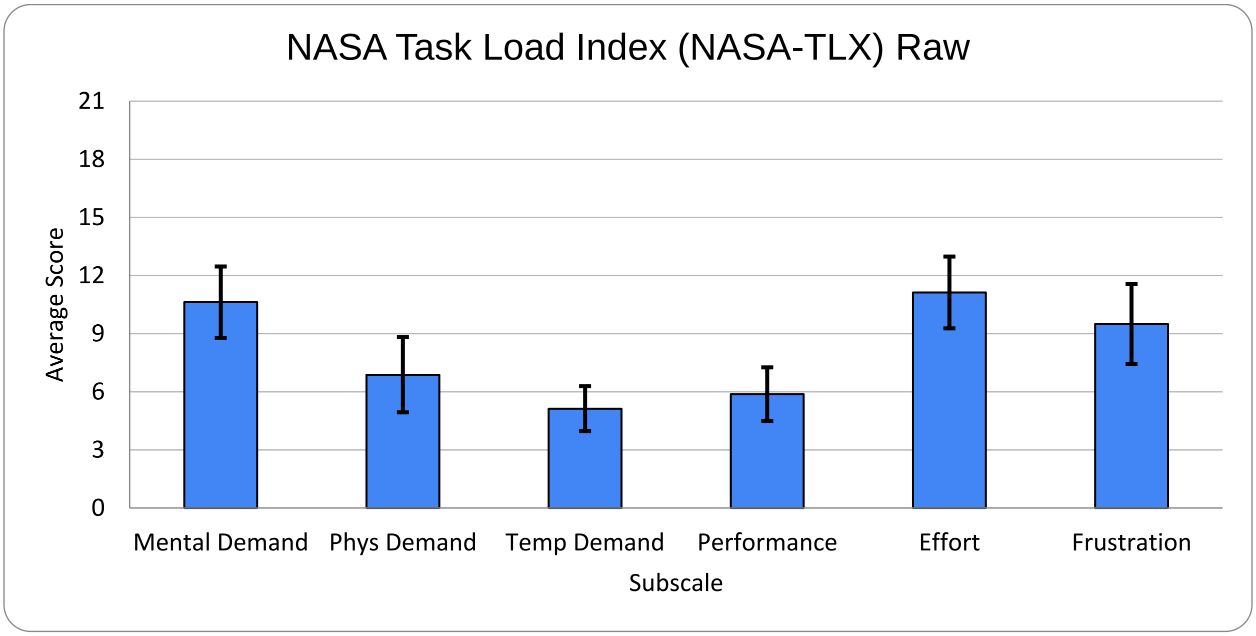
Note: Image is edited from Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. Intl. Journal of Human-Computer Interaction, 24(6), 574-594. https://doi.org/10.1080/10447310802205776

NASA-Task Load Index (NASA-TLX) Raw

Three of the NASA-TLX subscales resulted in lower middle-range scores: mental demand (M = 10.63, SD = 5.21), effort (M = 11.13, SD = 5.25), and frustration (M = 9.5, SD = 5.83). The other three subscales resulted in low scores: physical demand (M = 6.88, SD = 5.49), temporal demand (M = 5.13, SD = 3.27), and performance (M = 5.88, SD = 3.91). This is shown in Figure H5.

Figure H 5

Epson Moverio BT-300: NASA-TLX Raw Scores



Note: Error bars represent the standard error (SE).

Simulator Sickness Questionnaire (SSQ)

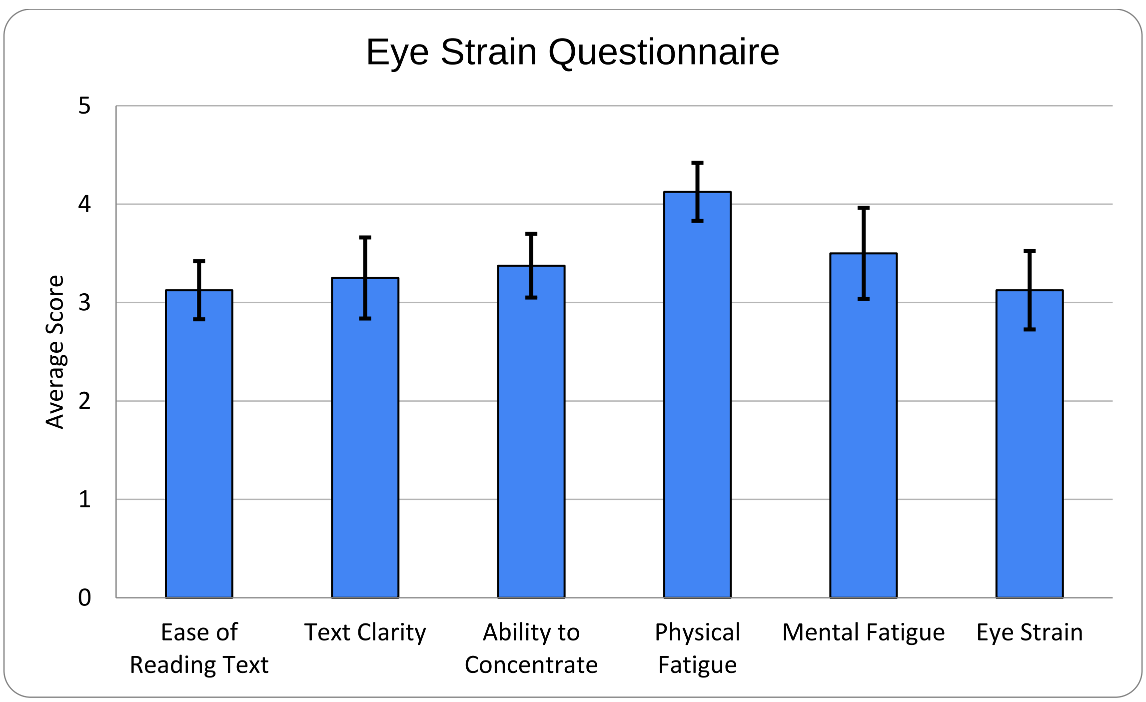
Total SSQ scores were in the “Extreme” level of severity (M = 42.54, SD = 41.69) (Hale & Stanney, 2014). The symptom profile followed a O > D > N pattern, with weighted oculomotor scores averaging the highest (M = 45.48, SD = 40.11), followed by weighted disorientation scores (M = 41.76, SD = 42.1), and weighted nausea scores (M = 22.66, SD = 28.36).

Eye Strain Questionnaire

These were coded so the most negative rating was 1 and most positive rating was 5. These included: ease of reading text (1 = very difficult, 5 = very easy), text clarity (1 = very dissatisfied, 5 = very satisfied), ability to concentrate (1 = very low, 5 = very high), physical fatigue (1 = very high, 5 = very low), mental fatigue (1 = very high, 5 = very low), and level of eyestrain (1 = very high, 5 = very low). The following items on the eye strain questionnaire resulted in middle-level scores: ease of reading text (M = 3.13, SD = 0.84), text clarity (M = 3.25, SD = 1.17), eye strain (M = 3.13, SD = 1.23). The following items resulted in higher scores, which meant participants felt more positive about their experience with the following: ability to concentrate (M = 3.38, SD = 0.92), physical fatigue (M = 4.13, SD = 0.84), and mental fatigue (M = 3.5, SD = 1.31). These average scores are shown in Figure H6.

Figure H 6

Epson Moverio BT-300: Eye Strain Questionnaire Scores



Note: Error bars represent the standard error (SE).

Magic Leap 1 & Wayfair Spaces

Task Success and Difficulty

All participants successfully completed seven out of the ten tasks. This is shown in Figure H7. The two tasks that participants struggled the most with were picking an object from another scene and picking an object from the items tab. These were also rated some of the most difficult tasks by participants: pick an object from another scene (M = 3.33, SD = 1.23); pick an object from the items tab (M = 3.78, SD = 1.39).The tutorial was also a difficult task to complete (M = 3.33, SD = 1.32). All of the task difficulty average scores are shown in Figure H8. Overall, the device was slightly easier to use (M = 4.22, SD = 0.83) than participants expected it to be at the beginning of the study (M = 3.78, SD = 0.44). This comparison is shown in Figure H9.

Figure H 7

Magic Leap 1 & Wayfair Spaces: Task Success Rate

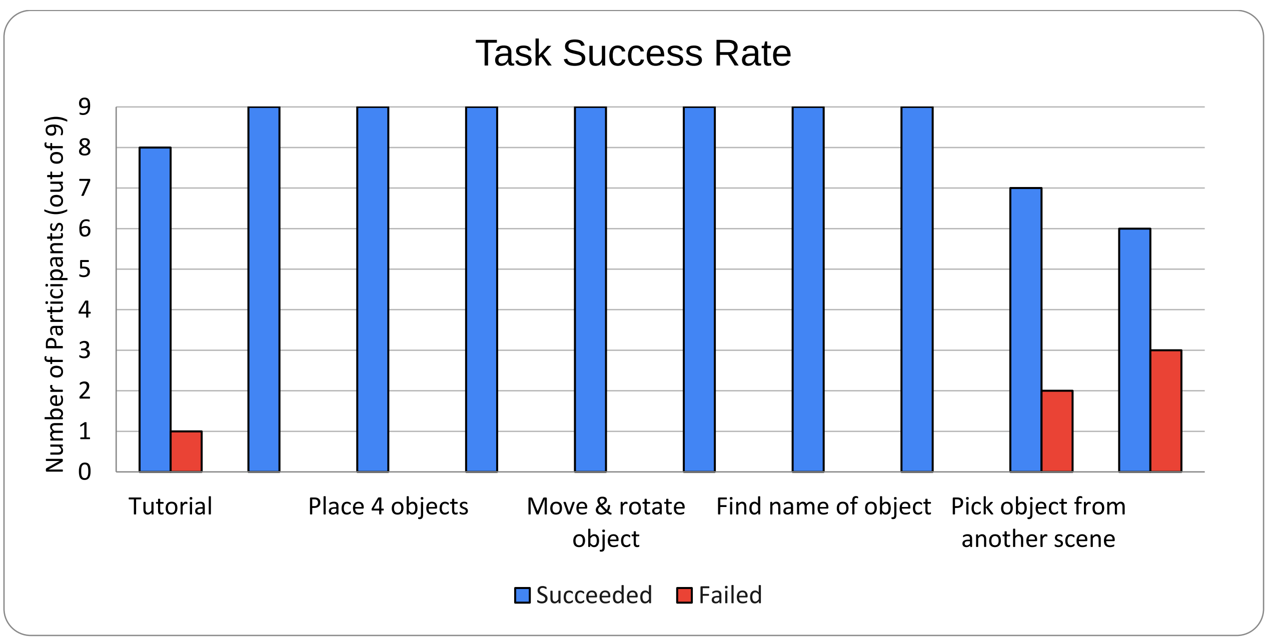
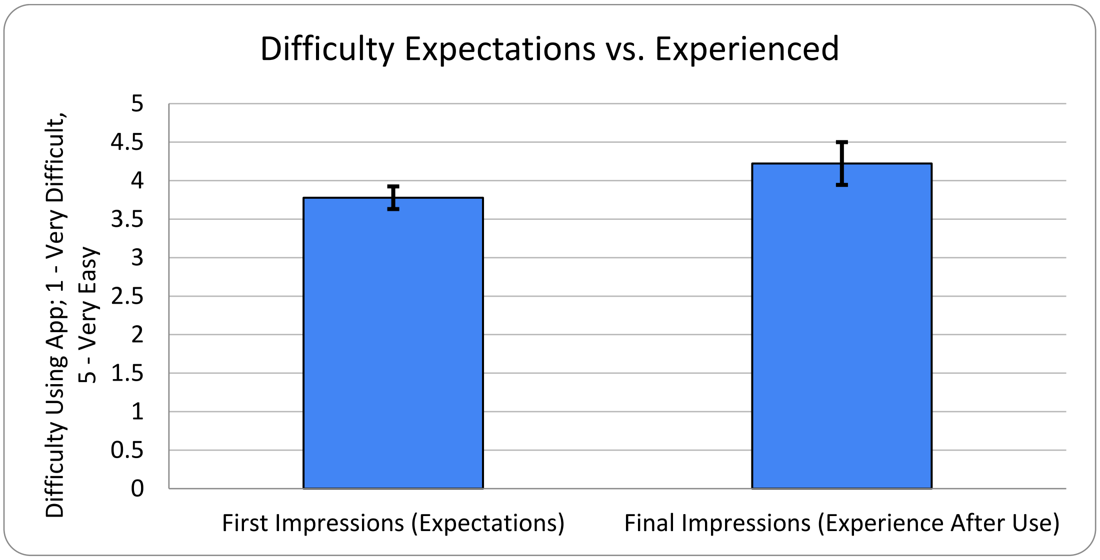


Figure H 8

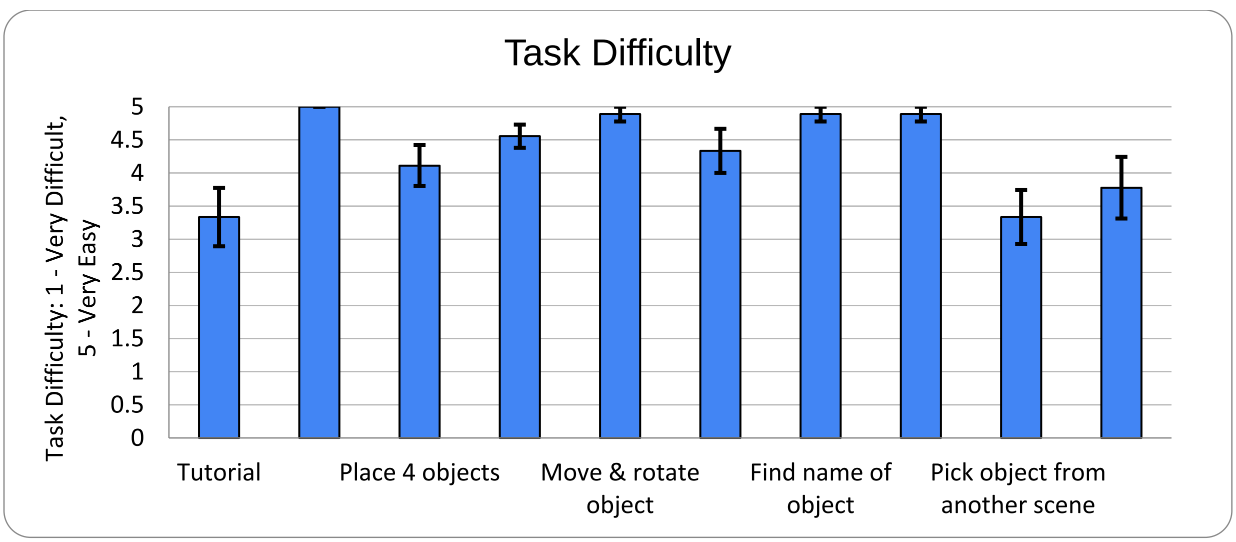
Magic Leap 1 & Wayfair Spaces: Device Difficulty Ratings Expectations vs. Actual Experience



Note: Error bars represent the standard error (SE).

Figure H 9

Magic Leap 1 & Wayfair Spaces: Difficulty Ratings by Task



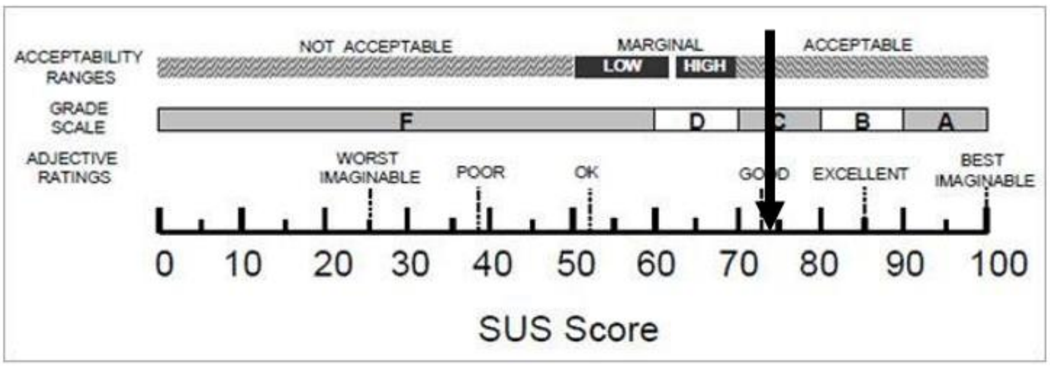
Note: Error bars represent the standard error (SE).

System Usability Scale (SUS)

SUS scores averaged to an "excellent" rating *(M* = 72.78, *SD* = 18.22). This means that participants rated the overall system as an excellent experience to use. This is shown in Figure H10.

Figure H 10

Magic Leap 1 & Wayfair Spaces: System Usability Scale Scores



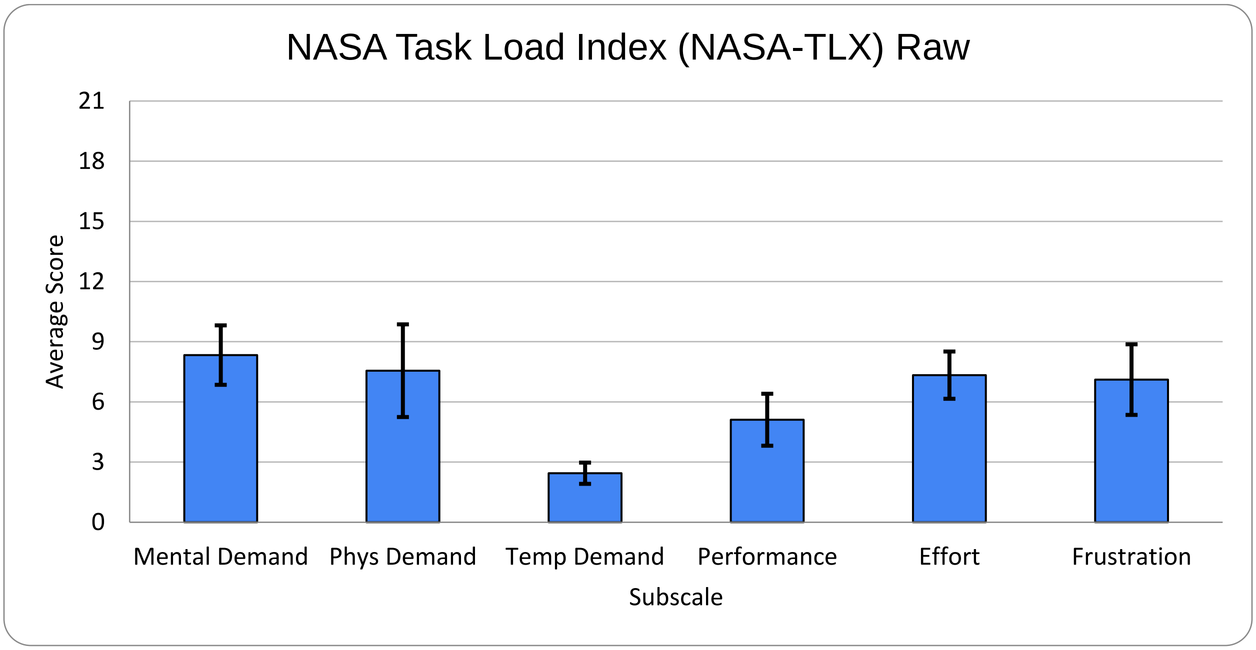
Note: Image is edited from Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. Intl. Journal of Human-Computer Interaction, 24(6), 574-594. https://doi.org/10.1080/10447310802205776

NASA-Task Load Index (NASA-TLX) Raw

All of the NASA-TLX subscales resulted in low scores: mental demand (*M* = 8.33 *SD* = 4.44), physical demand (*M* = 7.56, *SD* = 6.93), temporal demand (*M* = 2.44, *SD* = 1.59), performance (*M* = 5.11, *SD* = 3.89), effort (*M* = 7.33, *SD* = 3.54), and frustration (*M* = 7.11, *SD* = 5.28). These are shown in Figure H11.

Figure H 11

Magic Leap 1 & Wayfair Spaces: NASA-TLX Raw Scores



Note: Error bars represent the standard error (SE).

Simulator Sickness Questionnaire (SSQ)

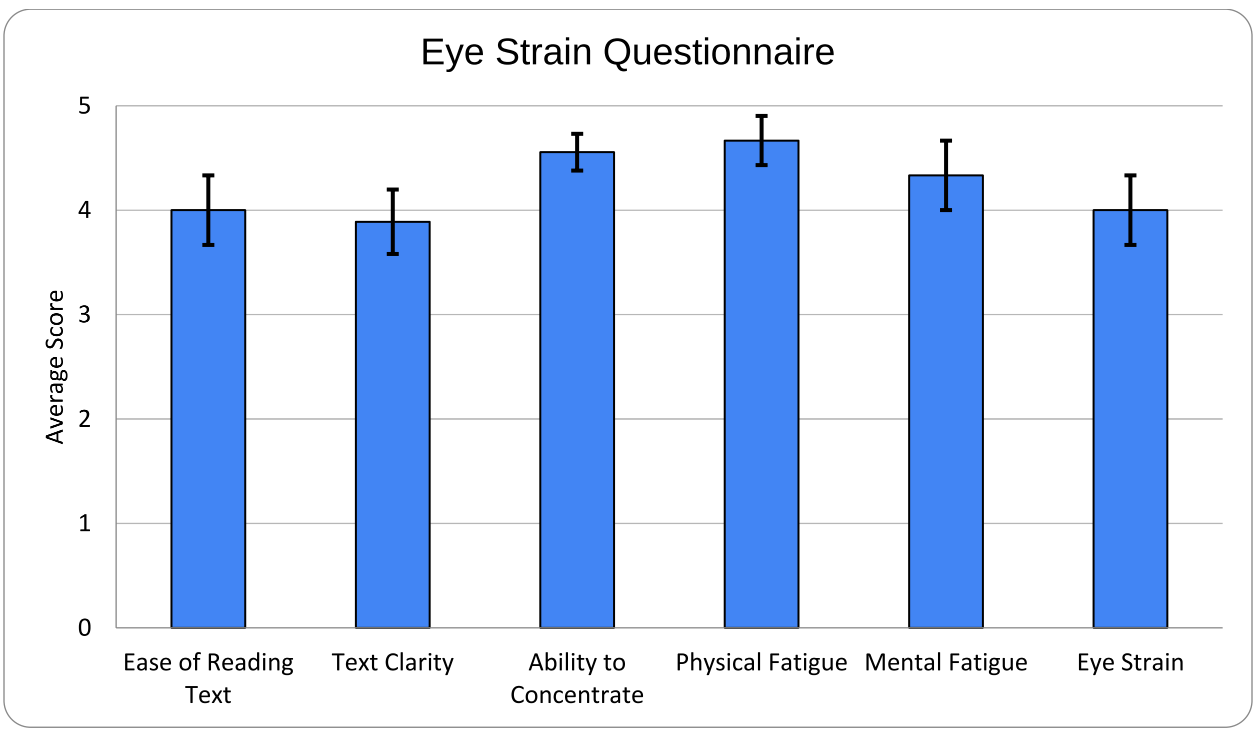
Total SSQ scores were in the "Moderate" level of severity (M = 19.95, SD = 22.13) (Hale & Stanney, 2014). The symptom profile followed a O > D > N pattern, with weighted oculomotor scores averaging the highest (M = 21.1, SD = 19.98), followed by weighted disorientation scores (M = 15.47, SD = 32.23), and weighted nausea scores (M = 13.78, SD = 11.79).

Eye Strain Questionnaire

Six questions 5-point Likert-scale questions were asked to assess a participant's level of eye strain. These were coded so the most negative rating was 1 and most positive rating was 5. These included: ease of reading text (1 = very difficult, 5 = very easy), text clarity (1 = very dissatisfied, 5 = very satisfied), ability to concentrate (1 = very low, 5 = very high), physical fatigue (1 = very high, 5 = very low), mental fatigue (1 = very high, 5 = very low), and level of eyestrain (1 = very high, 5 = very low). All of the items on the eye strain questionnaire resulted in high scores, which meant participants felt more positive about their experience with the following: ease of reading text (M = 4, SD = 1), text clarity (M =3.89, SD = 0.93), ability to concentrate (M = 4.56, SD = 0.53), physical fatigue (M = 4.67, SD = 0.71), mental fatigue (M = 4.33, SD = 1), and eye strain (M = 4, SD = 1). These average scores are shown in Figure H12.

Figure H 12

Magic Leap 1 & Wayfair Spaces: Eye Strain Questionnaire Scores



Note: Error bars represent the standard error (SE).

Magic Leap 2 & Fan Blade Replacement Application

Task Success and Difficulty

All participants successfully completed seven out of the eleven tasks. This is shown in Figure H13. The tasks that participants struggled the most with were removing the fan blade, replace the rear cover, and replace the front cover. These were also rated the most difficult task by participants: remove the fan blade *(M* = 2.29, *SD* = 1.11), replace the rear cover *(M* = 2.57*, SD* = 0.98), replace the front cover *(M* = 2.57, *SD* = 0.98)*.* Participants stated that they needed more virtual instruction in order to understand what to do with the physical object (e.g., add an animation that shows them how to take off the front cover). The difficult ratings by task are shown in Figure H14. Overall, the device was about as difficult to use *(M* = 3.71, *SD* = 0.76) as participants expected it to be at the beginning of the study *(M* = 4, *SD* = 0.82)*.* This comparison is shown in Figure H15.

Figure H 13

Magic Leap 2 & Fan Blade Replacement: Task Success Rate

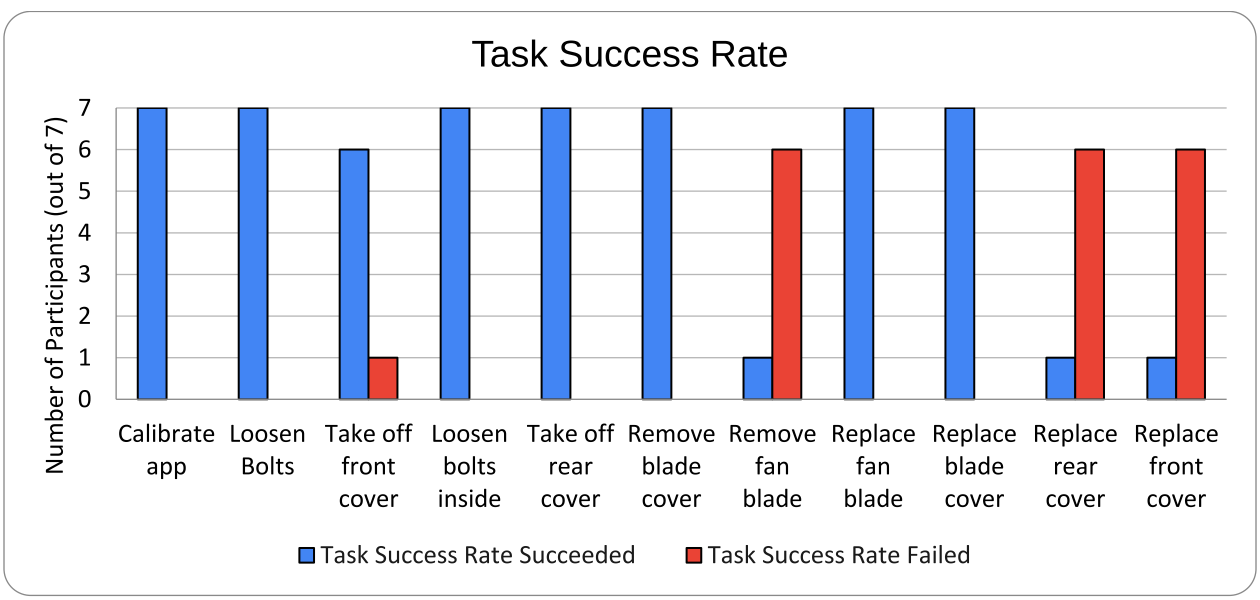
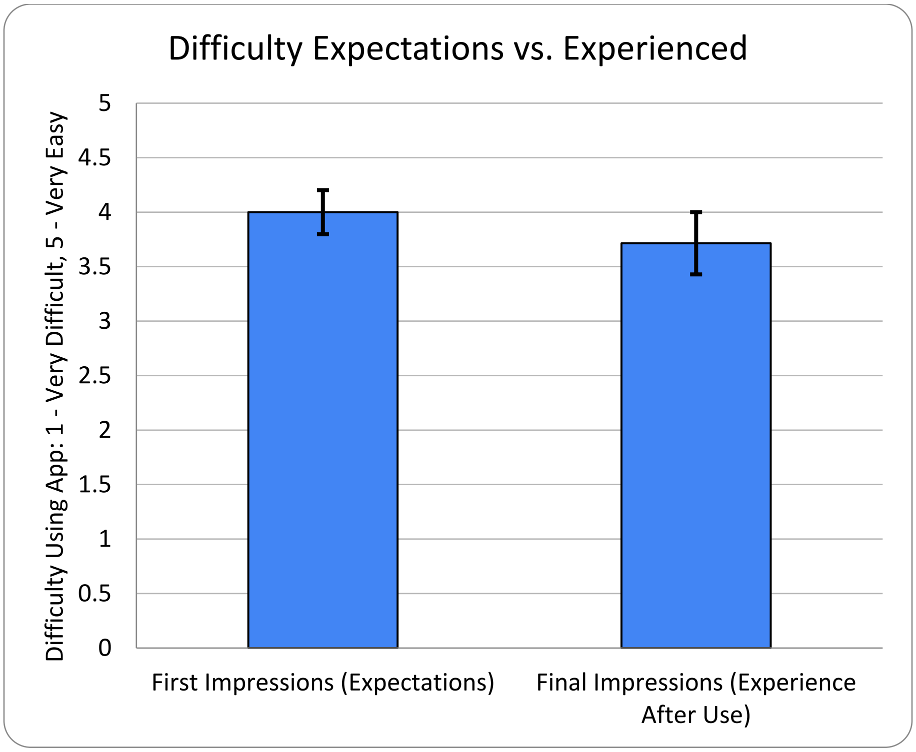


Figure H 14

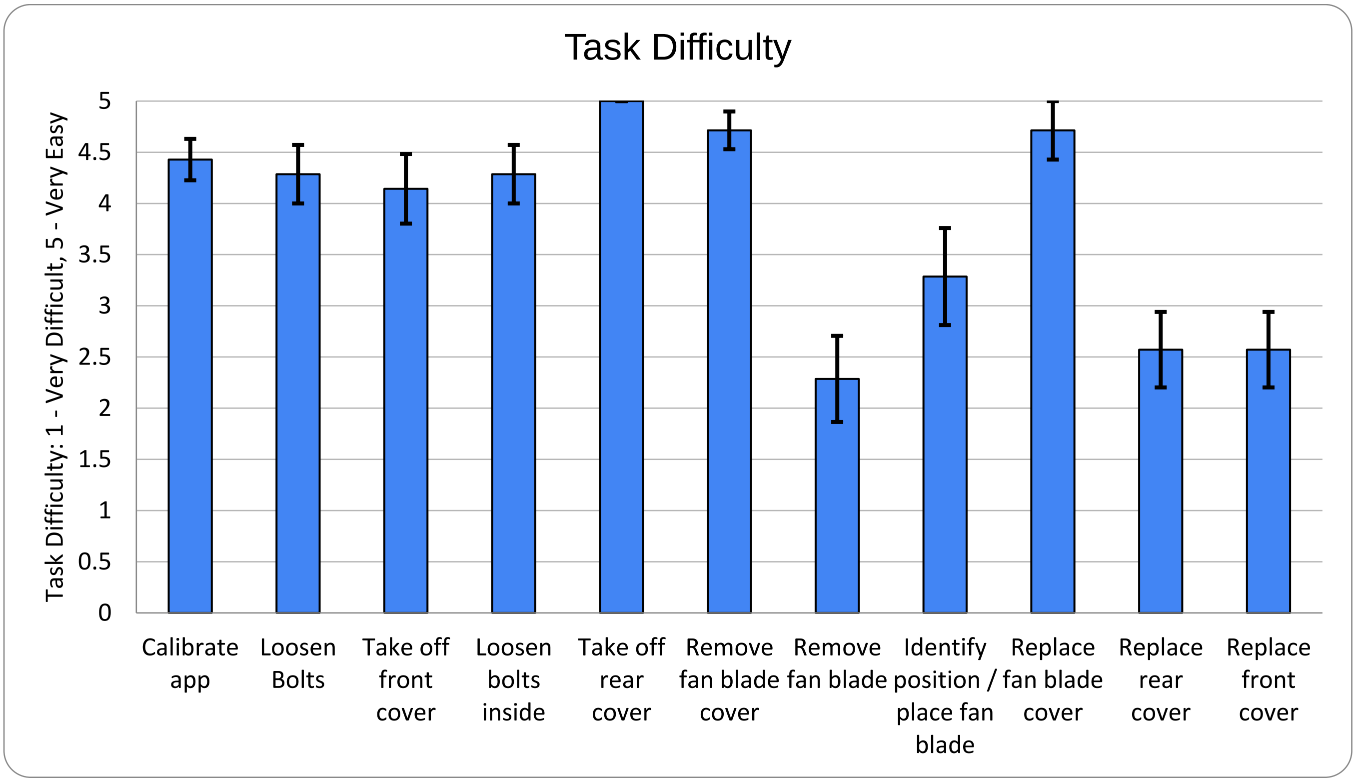
Magic Leap 2 & Fan Blade Replacement: Device Difficulty Ratings Expectations vs. Actual Experience



Note: Error bars represent the standard error (SE).

Figure H 15

Magic Leap 2 & Fan Blade Replacement: Difficulty Ratings by Task



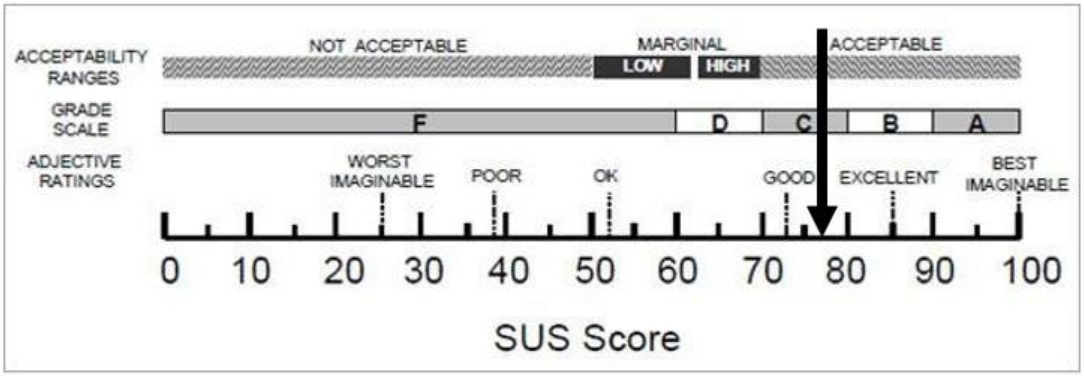
Note: Error bars represent the standard error (SE).

System Usability Scale (SUS)

SUS scores averaged to an "excellent" rating *(M* = 76.43, *SD* = 15.74)*.* This means that participants rated the overall system as an excellent experience to use. This is shown in Figure H16.

Figure H 16

Magic Leap 2 & Fan Blade Replacement: System Usability Scale Scores



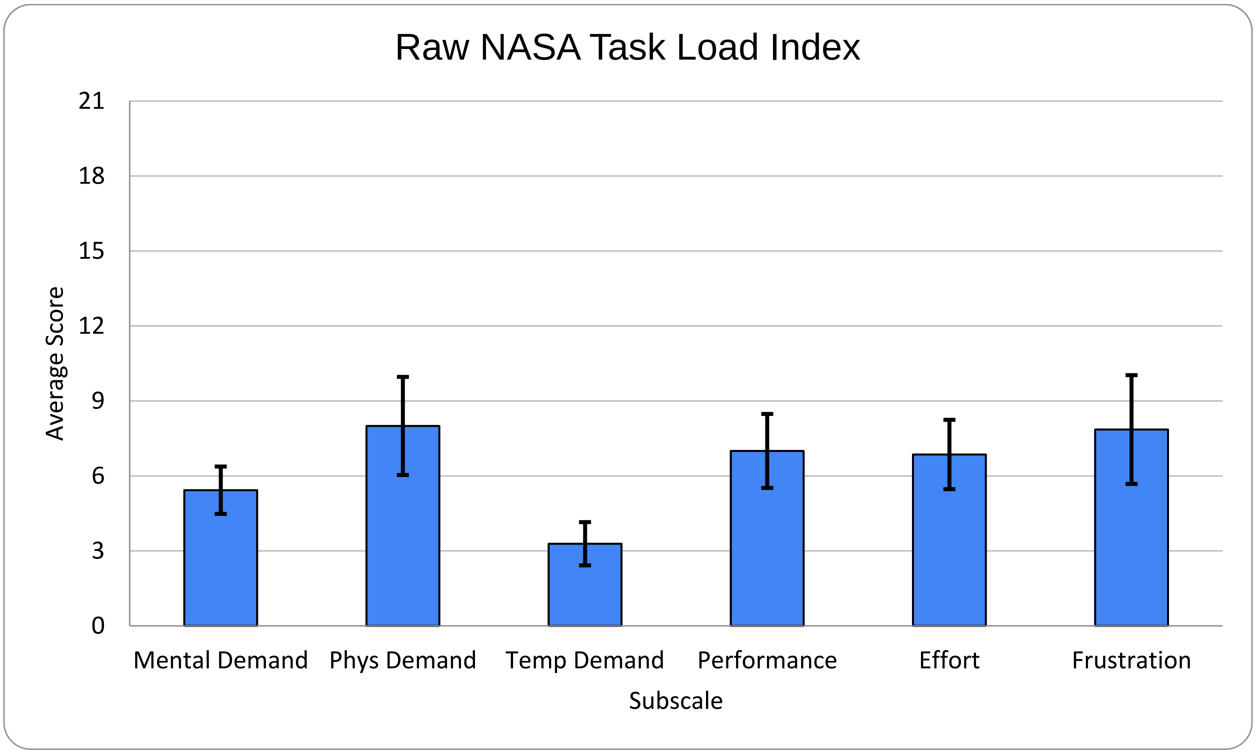
Note: Image is edited from Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. Intl. Journal of Human-Computer Interaction, 24(6), 574-594. https://doi.org/10.1080/10447310802205776

NASA-Task Load Index (NASA-TLX) Raw

All of the NASA-TLX subscales resulted in low scores: mental demand (*M* = 5.43, *SD* = 2.51), physical demand (*M* = 8, *SD* = 5.2), temporal demand (*M* = 3.29, *SD* = 2.29), performance (*M* = 7, *SD* = 3.92), effort (*M* = 6.86, *SD* = 3.67), and frustration (*M* = 7.86, *SD* = 5.76). This is shown in Figure H17.

Figure H 17

Magic Leap 2 & Fan Blade Replacement: NASA-TLX Raw Scores



Note: Error bars represent the standard error (SE).

Simulator Sickness Questionnaire (SSQ)

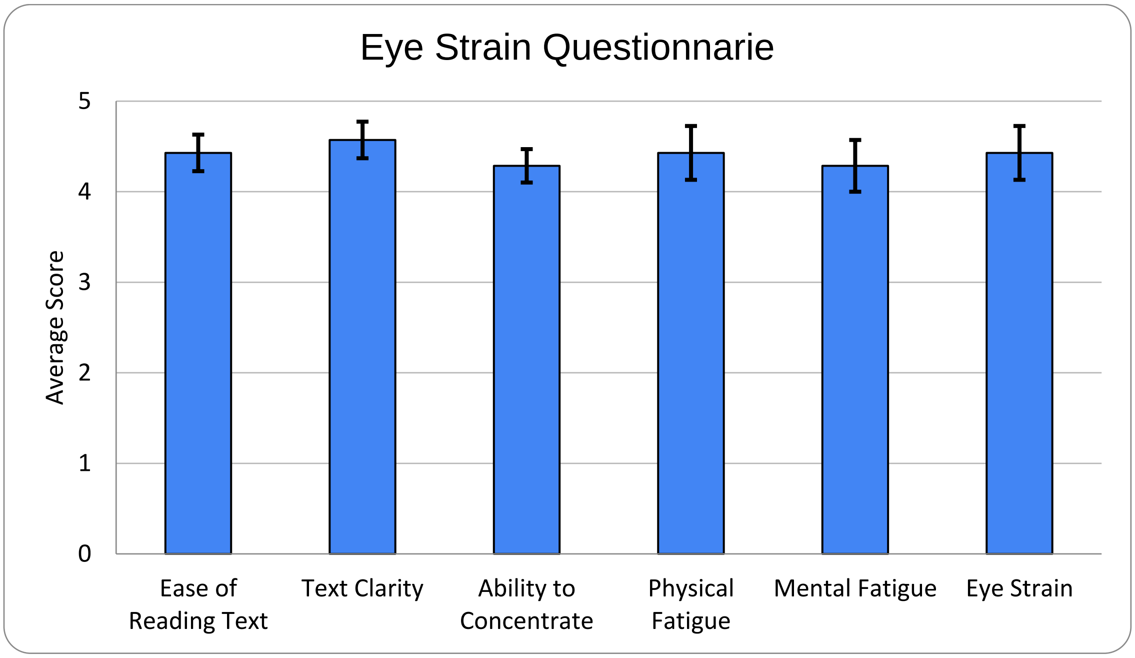
Total SSQ scores were in the "Low" level of severity (M = 12.82, SD = 11.2) (Hale & Stanney, 2014). The symptom profile followed a O > N > D pattern, with weighted oculomotor scores averaging the highest (M = 15.16, SD = 13.84), followed by weighted nausea scores (M = 10.9, SD = 10.2), and weighted disorientation scores (M = 3.98, SD = 10.52).

Eye Strain Questionnaire

Six questions 5-point Likert-scale questions were asked to assess a participant's level of eye strain. These were coded so the most negative rating was 1 and most positive rating was 5. These included: ease of reading text (1 = very difficult, 5 = very easy), text clarity (1 = very dissatisfied, 5 = very satisfied), ability to concentrate (1 = very low, 5 = very high), physical fatigue (1 = very high, 5 = very low), mental fatigue (1 = very high, 5 = very low), and level of eyestrain (1 = very high, 5 = very low). All of the items on the eye strain questionnaire resulted in high scores, which meant participants felt more positive about their experience with the following: ease of reading text (M = 4.43, SD = 0.54), text clarity (M = 4.57, SD = 0.54), ability to concentrate (M = 4.29, SD = 0.49), physical fatigue (M = 4.43, SD = 0.79), mental fatigue (M = 4.29, SD = 0.76), and eye strain (M = 4.43, SD = 0.79). This is shown in Figure H18.

Figure H 18

Magic Leap 2 & Fan Blade Replacement: Eye Strain Questionnaire Scores



Note: Error bars represent the standard error (SE).

Meta Quest Pro & ShapesXR

Task Success and Difficulty

All eight participants successfully completed four out of the ten tasks. This is shown in Figure H19. Tasks that participants struggled with the most were changing the colors of virtual objects, taking pictures of the final prototype, and placing objects. Even though participants struggled with completing these tasks correctly, no task was rated, on average, a 3 on the difficulty scale. This is because participants stated that it was difficult to remember how to begin these tasks, but once they were instructed how to do it, the task itself was easy. The most difficult tasks were creating the office space overall, (M = 3.63, SD = 0.52), placing objects (M = 3.88, SD = 0.64), placing an image in the space (M = 3.88, SD =

0.64), and taking photos of the finished prototype *(M* = 3.63, *SD* = 1.06)*.* All average task difficulty scores are shown in Figure H20. Overall, the device was about as difficult to use *(M* = 3.5, *SD* = 0.54) as participants expected it to be at the beginning of the study *(M* = 3.63, *SD* = 0.92)*.* This comparison is shown in Figure H21.

Figure H 19

Meta Quest Pro & ShapesXR: Task Success Rate

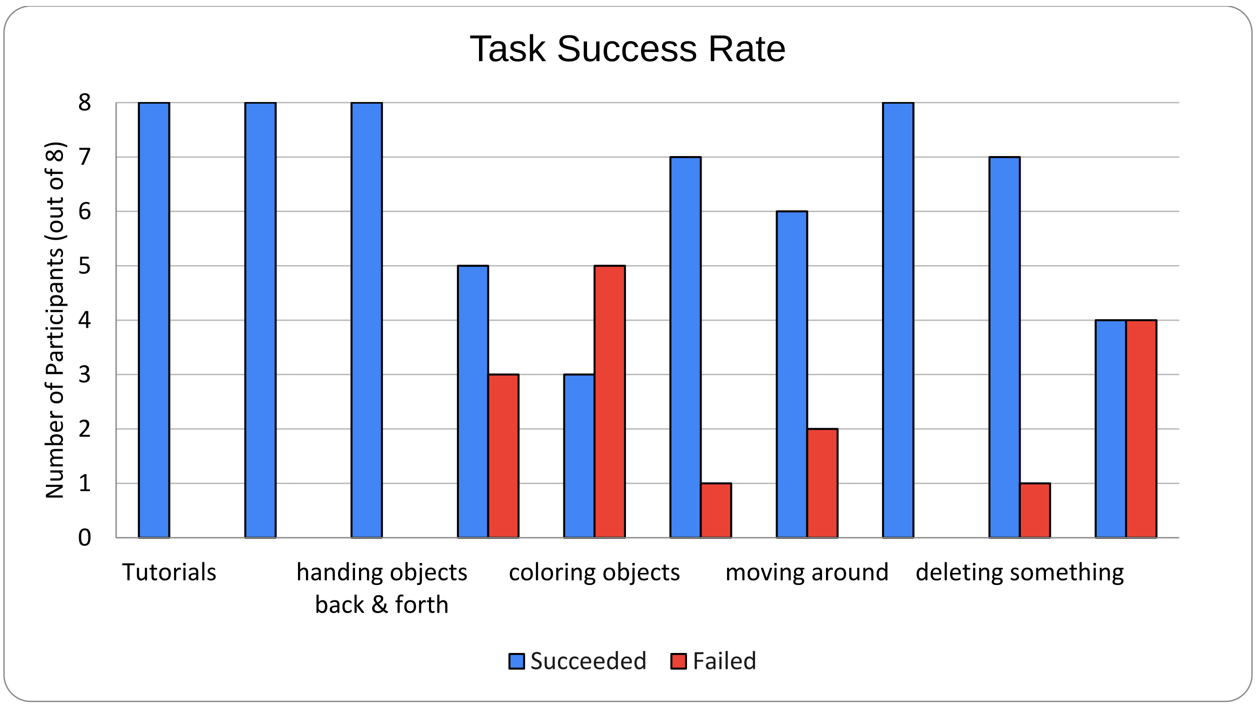
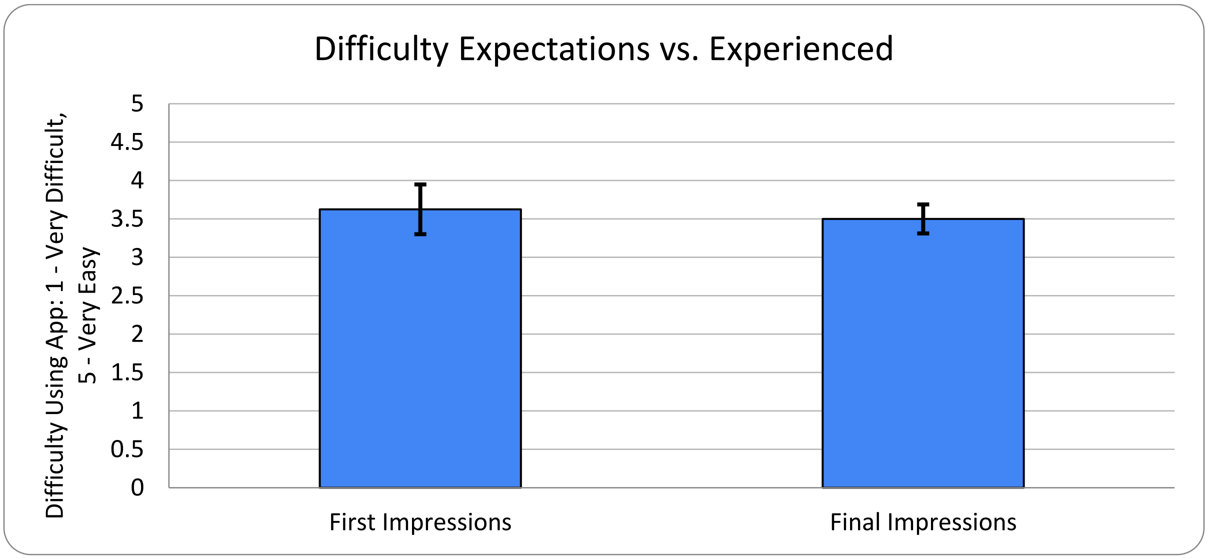


Figure H 20

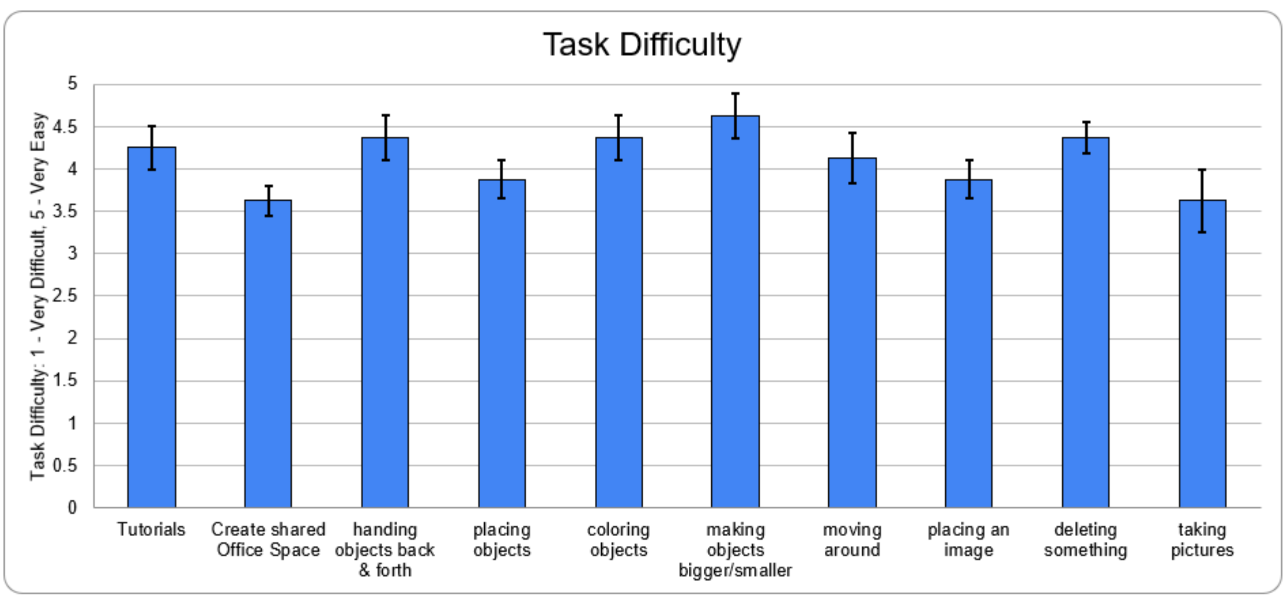
Meta Quest Pro & ShapesXR: Device Difficulty Ratings Expectations vs. Actual Experience



Note: Error bars represent the standard error (SE).

Figure H 21

Meta Quest Pro & ShapesXR: Difficulty Ratings by Task



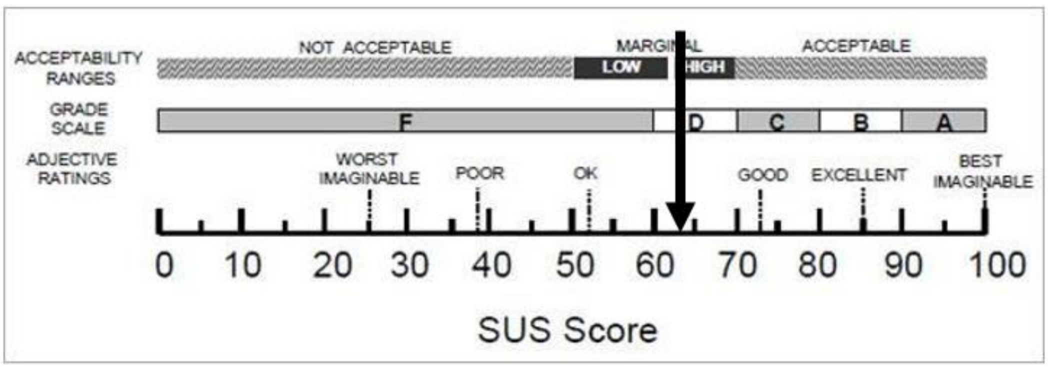
Note: Error bars represent the standard error (SE).

System Usability Scale (SUS)

SUS scores averaged to a "good" rating *(M* = 63.75, *SD* = 18.27)*.* This means that participants reported having a good experience with the system as a whole. Figure H22 shows how this score fairs on the SUS.

Figure H 22

Meta Quest Pro & ShapesXR: System Usability Scale Scores



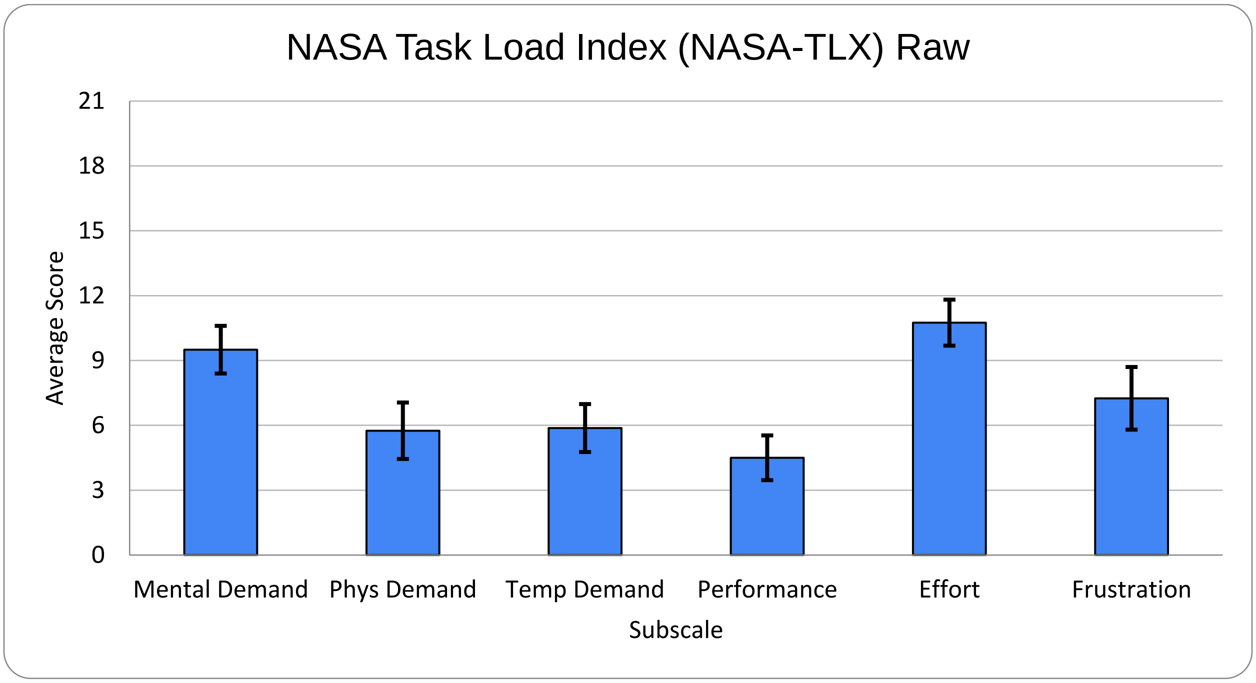
Note: Image is edited from Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. Intl. Journal of Human-Computer Interaction, 24(6), 574-594. https://doi.org/10.1080/10447310802205776

NASA-Task Load Index (NASA-TLX) Raw

Two of the NASA-TLX subscales resulted in lower middle-range scores: mental demand (*M* = 9.5, *SD* = 3.12), and effort (*M* = 10.75, *SD* = 3.01). The other subscales resulted in low scores: physical demand (*M* = 5.75, *SD* = 3.69), temporal demand (*M* = 5.88, *SD* = 3.14), performance (*M* = 4.5, *SD* = 2.93), and frustration (*M* = 7.25, *SD* = 4.1). This is shown in Figure H23.

Figure H 23

Meta Quest Pro & ShapesXR: NASA-TLX Raw Scores



Note: Error bars represent the standard error (SE).

Simulator Sickness Questionnaire (SSQ)

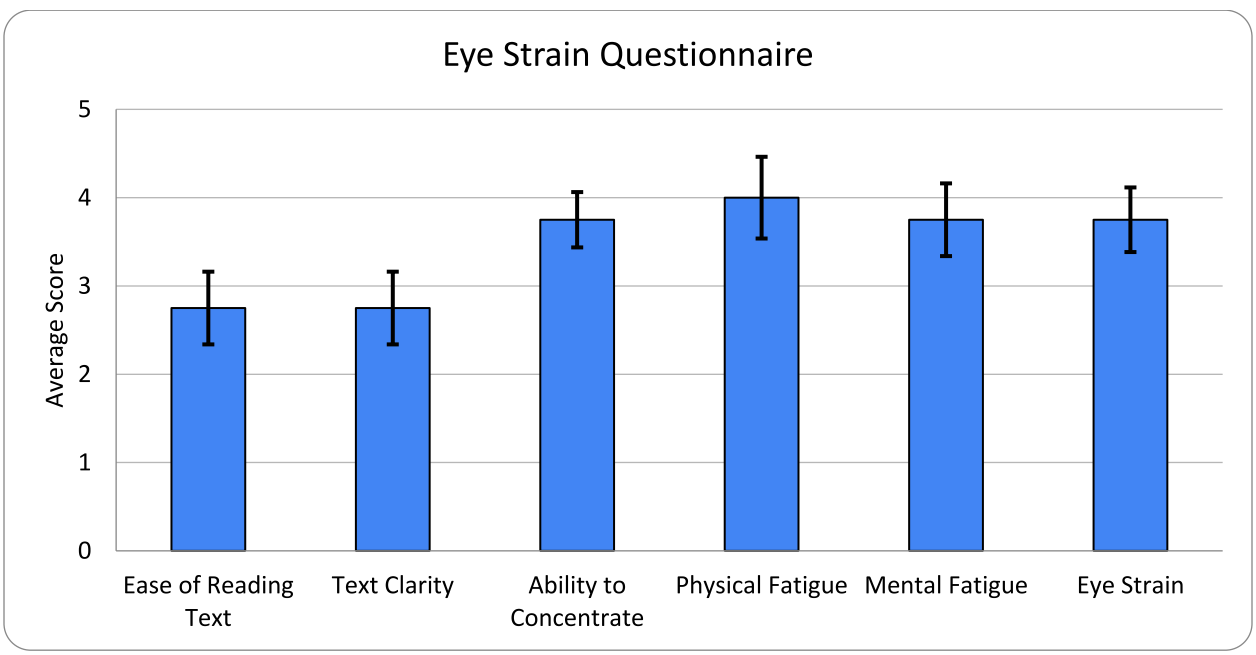
Total SSQ scores were in the "Medium" level of severity (*M* = 27.12, *SD* = 16.21) (Hale & Stanney, 2014). The symptom profile followed a O > N > D pattern, with weighted oculomotor scores averaging the highest (*M* = 32.22, *SD* = 19.75), followed by weighted nausea scores (*M* = 17.89, *SD* = 7.96), and weighted disorientation scores (*M* = 15.66, *SD* = 20.29).

Eye Strain Questionnaire

Six questions 5-point Likert-scale questions were asked to assess a participant's level of eye strain. These were coded so the most negative rating was 1 and most positive rating was 5. These included: ease of reading text (1 = very difficult, 5 = very easy), text clarity (1 = very dissatisfied, 5 = very satisfied), ability to concentrate (1 = very low, 5 = very high), physical fatigue (1 = very high, 5 = very low), mental fatigue (1 = very high, 5 = very low), and level of eyestrain (1 = very high, 5 = very low). The following items on the eye strain questionnaire resulted in middle-level scores: ease of reading text (*M* = 2.75, *SD* = 1.17), text clarity (*M* = 2.75, *SD* = 1.17). The following items resulted in higher scores, which meant participants felt more positive about their experience with the following: ability to concentrate (*M* = 3.75, *SD* = 0.89), physical fatigue (*M* = 4, *SD* = 4.5), mental fatigue (*M* = 3.75, *SD* = 1.17), and eye strain (*M* = 3.75, *SD* = 1.04). These scores are shown in Figure H24.

Figure H 24

Meta Quest Pro & ShapesXR: Eye Strain Questionnaire Scores



Note: Error bars represent the standard error (SE).

Mobile Phone & Google Maps Live View

Task Success and Difficulty

All nine participants successfully completed all tasks successfully. This is shown in Figure H25. Participants reported that the most difficult task to complete was starting the AR walking directions *(M* = 3.33, *SD* = 1.23). All other tasks were rated easy and can be seen in Figure H26. Overall, the device was as difficult to use *(M* = 4, *SD* = 0.87) as participants expected it to be at the beginning of the study *(M* = 4, *SD* = 0.87). This comparison is shown in Figure H28.

Figure H 25

Mobile Phone & Google Maps Live View: Task Success Rate

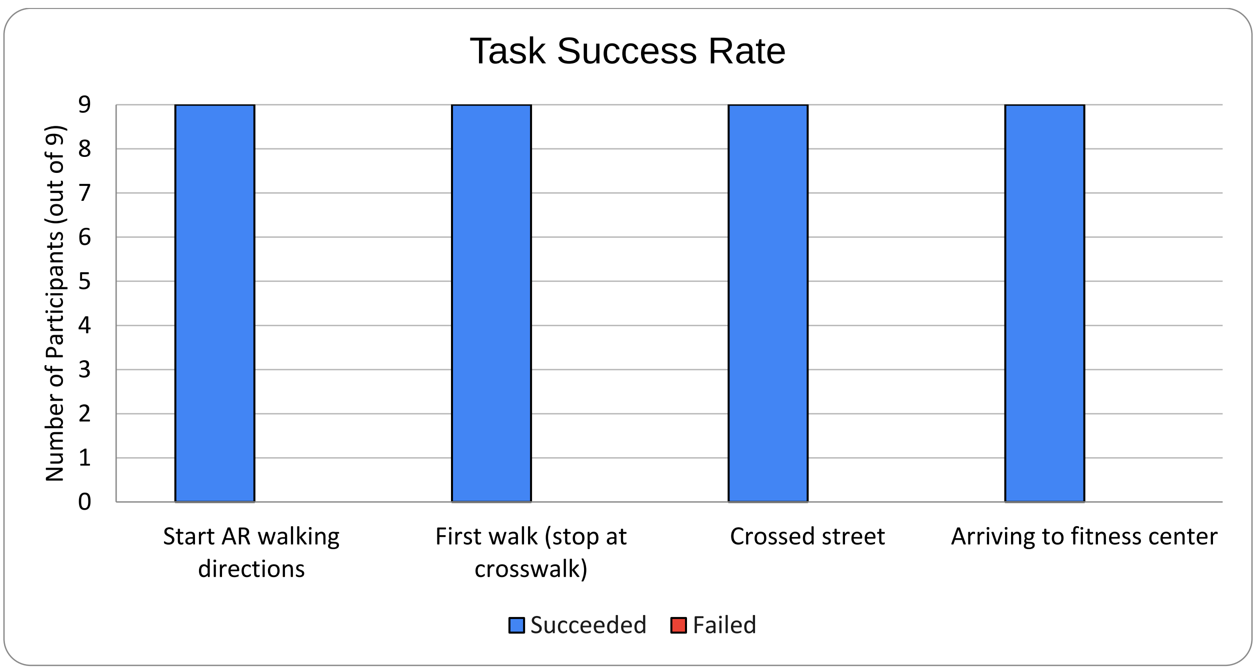
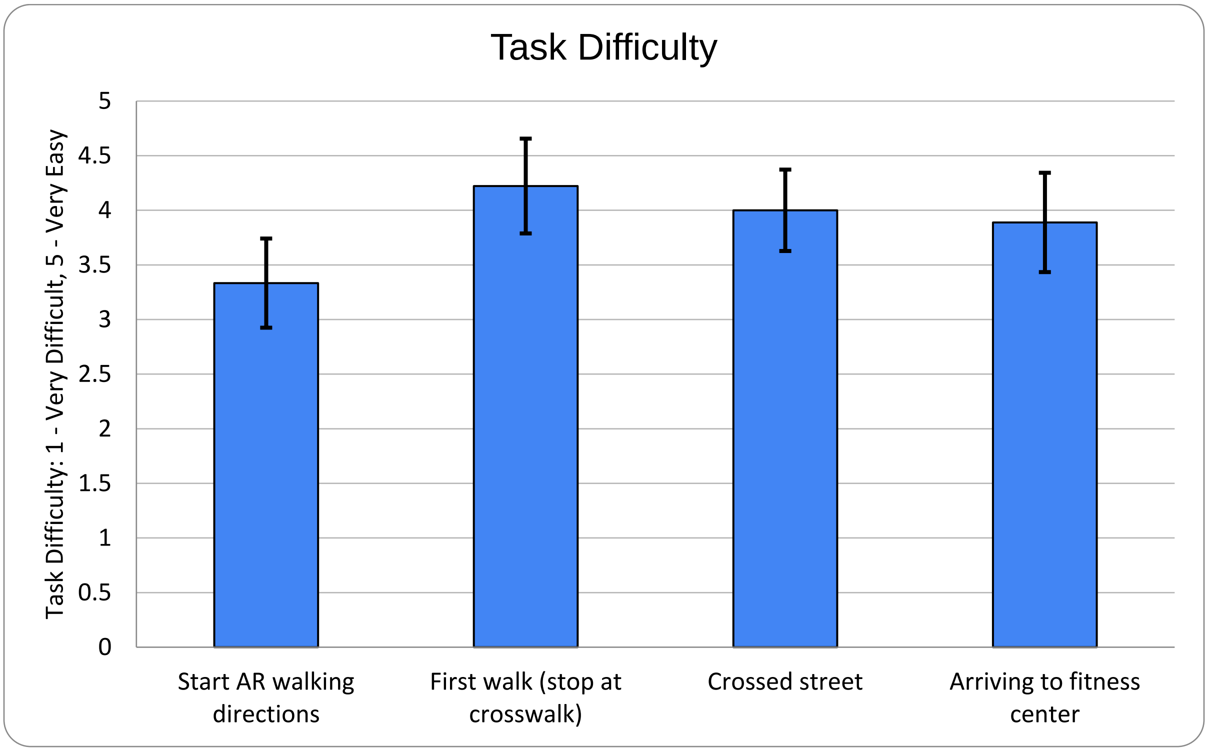


Figure H 26

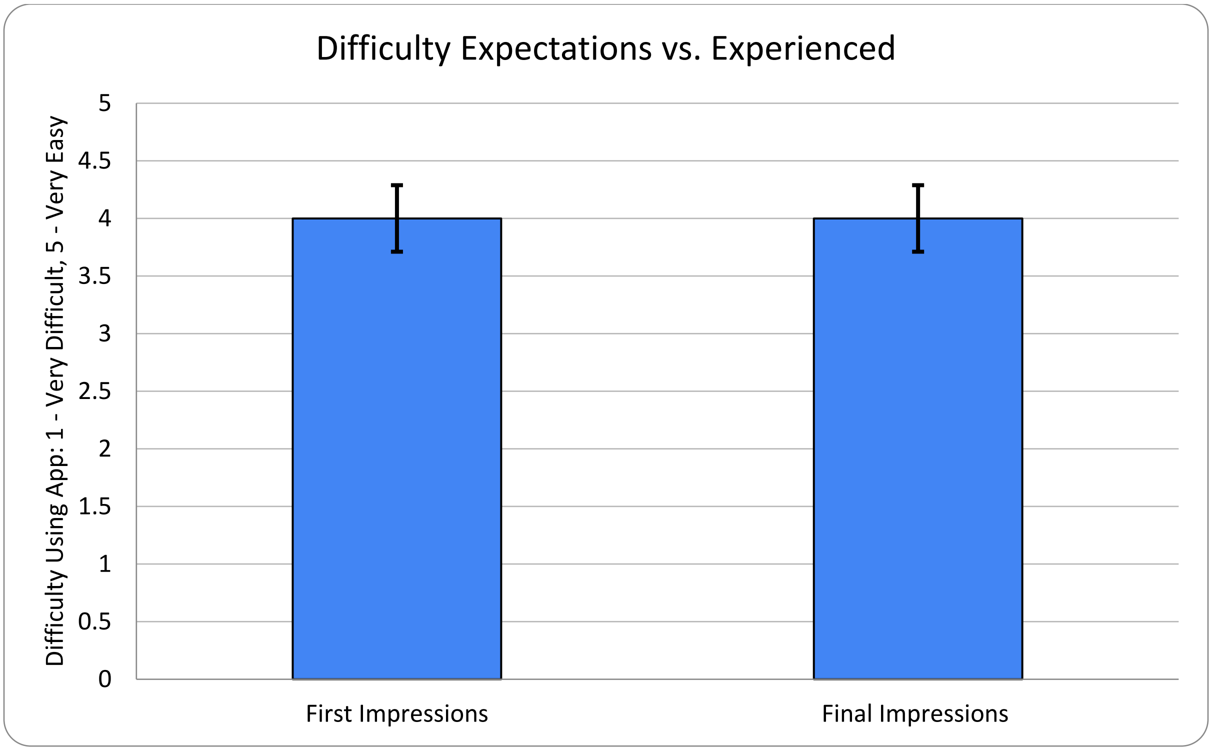
Mobile Phone & Google Maps Live View: Difficulty Ratings by Task



Note: Error bars represent the standard error (SE).

Figure H 27

Mobile Phone & Google Maps Live View: Device Difficulty Ratings Expectations vs. Actual Experience



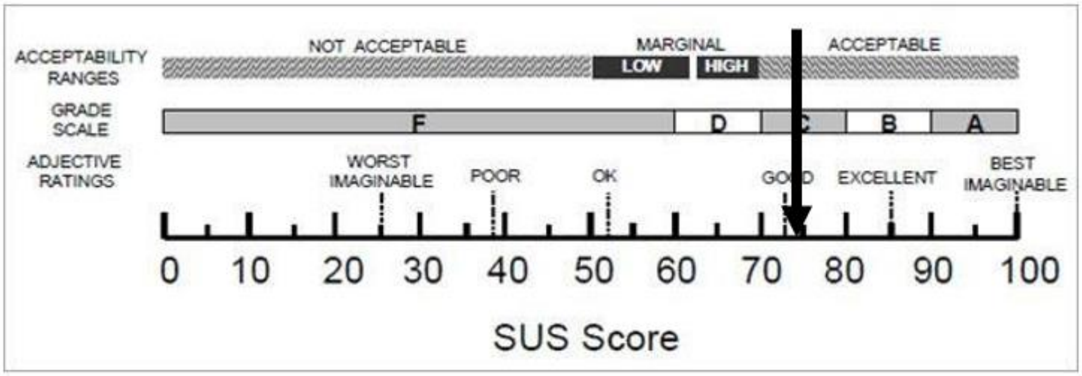
Note: Error bars represent the standard error (SE).

System Usability Scale (SUS)

SUS scores averaged to a "good" rating *(M* = 74.17, *SD* = 18.96)*.* This means that participants reported having a good experience with the system as a whole. Figure H28 shows how this score fairs on the SUS.

Figure H 28

Mobile Phone & Google Maps Live View: System Usability Scale Scores



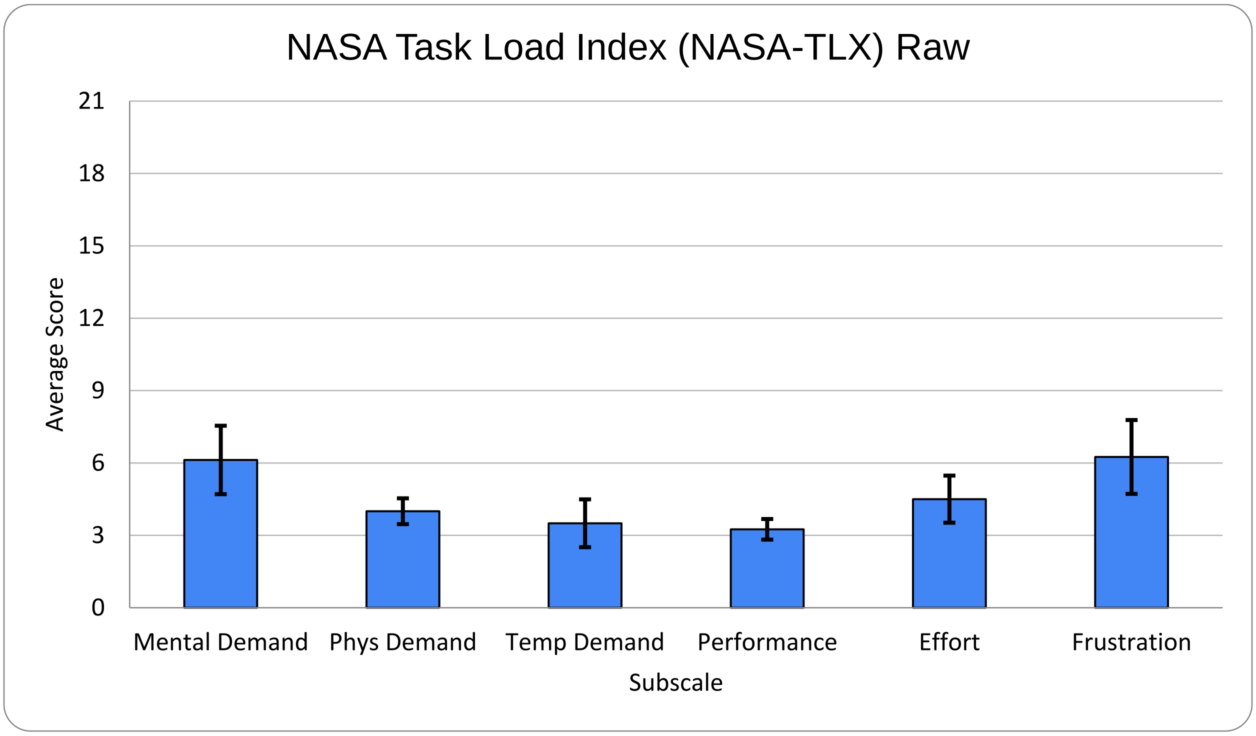
Note: Image is edited from Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. Intl. Journal of Human-Computer Interaction, 24(6), 574-594. https://doi.org/10.1080/10447310802205776

NASA-Task Load Index (NASA-TLX) Raw

All of the NASA-TLX subscales resulted in low scores: mental demand (M = 6.13, SD = 4.26), physical demand (M = 4, SD = 1.6), temporal demand (M = 3.5, SD = 2.98), performance (M = 3.25, SD = 1.28), effort (M = 4.5, SD = 2.93), and frustration (M = 6.25, SD = 4.6). These scores can be seen in Figure H29.

Figure H 29

Mobile Phone & Google Maps Live View: NASA-TLX Raw Scores



Note: Error bars represent the standard error (SE).

Simulator Sickness Questionnaire (SSQ)

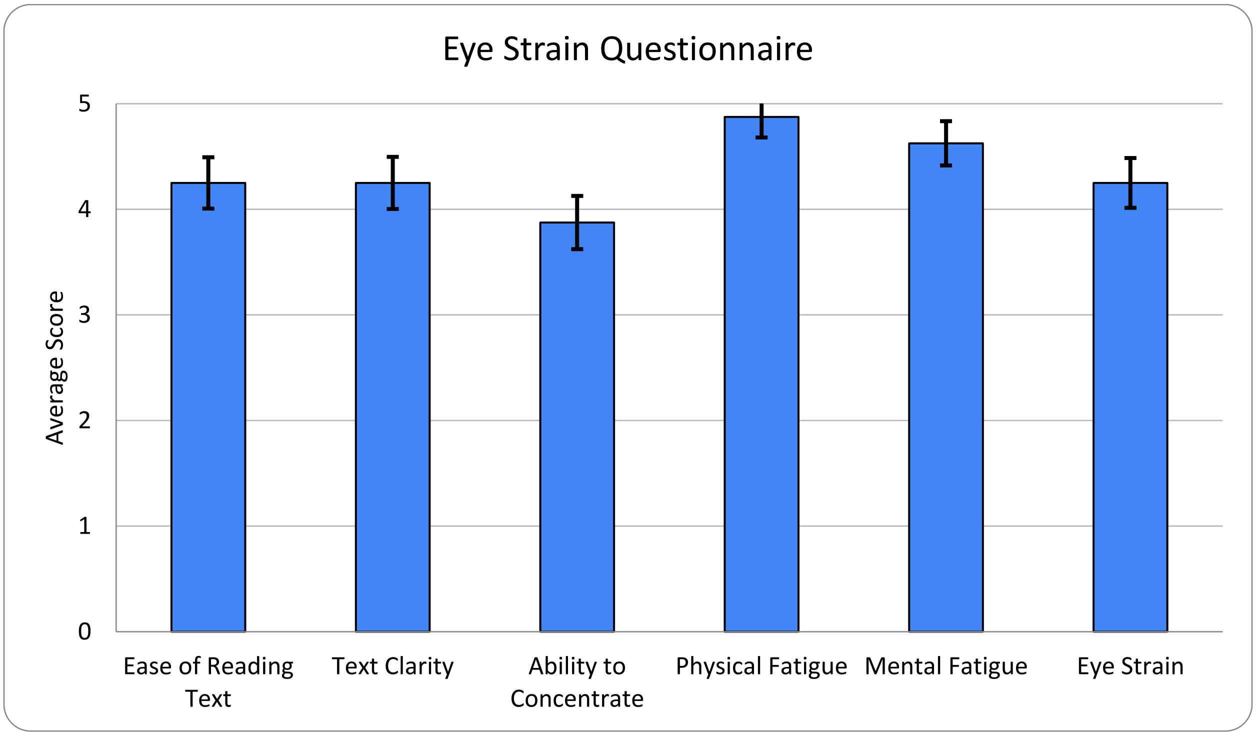
Total SSQ scores were in the "low" level of severity (M = 6.65, SD = 3.74) (Hale & Stanney, 2014). The symptom profile followed a O > N > D pattern, with weighted oculomotor scores averaging the highest (M = 7.58, SD = 5.36), followed by weighted nausea scores (M = 6.36, SD = 9.54), and weighted disorientation scores (M = 1.55, SD = 4.64).

Eye Strain Questionnaire

Six questions 5-point Likert-scale questions were asked to assess a participant's level of eye strain. These were coded so the most negative rating was 1 and most positive rating was 5. These included: ease of reading text (1 = very difficult, 5 = very easy), text clarity (1 = very dissatisfied, 5 = very satisfied), ability to concentrate (1 = very low, 5 = very high), physical fatigue (1 = very high, 5 = very low), mental fatigue (1 = very high, 5 = very low), and level of eyestrain (1 = very high, 5 = very low). All of the items on the eye strain questionnaire resulted in high scores, which meant participants felt more positive about their experience with the following: ease of reading text (M = 4.25, SD = 0.73), text clarity (M = 4.25, SD = 0.74), ability to concentrate (M = 3.88, SD = 0.76), physical fatigue (M = 4.88, SD = 0.58), mental fatigue (M = 4.63, SD = 0.63), and eye strain (M = 4.25, SD = 0.71). Figure H30 shows these scores.

Figure H 30

Mobile Phone & Google Maps Live View: Eye Strain Questionnaire Scores



Note: Error bars represent the standard error (SE).

Appendix I

The Revised 12 AR and MR Usability Heuristics After Step 8 was Completed

This Appendix gives a brief overview of each of the 12 heuristics in the Derby Dozen toolkit. Using the toolkit, an evaluator may rate each of these checklist items as a “yes”, “somewhat”, “no”, or “not applicable (N/A)” and provides a space for evaluators to give additional information about why they gave the item that rating.

Heuristic 1: Unboxing & Setting Up

1. Is the unboxing process a positive experience?
2. 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?
3. Is a quick start guide available with the device?
4. Is it easy to set up the device and/or application between uses?
5. Is a call to action (QR code, instructions to use AR, instructions to download an app, etc.) clearly marked in the physical space?

Heuristic 2: Instructions

6. Is there the option of a tutorial upon first use of the device and/or application?
7. Does the tutorial explain all of the necessary actions/mechanics to use the device and/or application?
8. Is the tutorial easy to understand for both novice and experienced users?
9. Are required interactions easy to learn?
10. Is help or documentation easily accessible for the application?
11. Are instructions easy to understand?

12. Do instructions provide actionable feedback?
13. If auditory instructions are given, do these instructions match what the user is seeing in the application?
14. Are error messages easy to understand?
15. Do error messages provide actionable feedback?
16. Does the device's user interface and/or application avoid irreversible errors?
17. Is there a way for the user to report errors or crashes to the developer?

Heuristic 3: Organization & Simplification

18. Is the user eased into the virtual environment?
19. Does the device's user interface and/or application avoid clutter, as appropriate?
20. Does the device's user interface and/or application avoid large amounts of text?
21. Does the screen space focus on the virtual elements rather than controls or other non-AR/MR features, as appropriate?
22. Is information organized in an understandable manner?
23. If the quantity of information is large, is it organized in a layered or hierarchical manner so it is easy to understand?
24. Does the device's user interface and/or application avoid tasks that involve a large amount of steps to complete?
25. Does the application make use of all of its AR/MR functions (including information that is visual, auditory, and involved other sensory modalities)?

Heuristic 4: Consistency

26. Are virtual elements easy to delete or close out of?

27. Can the user pause the application at any point?
28. Are all aspects of the device's user interface and/or application (virtual elements, controls, text, etc.) clear and readable?
29. Are virtual elements sized appropriately?
30. Are virtual elements rendered a reasonable distance away from the user's targeted point?
31. For mobile devices, are the controls based on known interactions for mobile devices?
32. For mobile devices, are landscape and portrait mode supported?
33. For mobile devices, is the application responsive?
34. Do virtual elements act as the user would expect them to in the real world?
35. Does the device and/or application avoid lag, delays, jitter, drift, and other forms of virtual element malfunctions?
36. Is the navigation consistent throughout the device and/or application?
37. Can the user navigate freely throughout aspects of the device and/or application?
38. Does the device allow for adjustment based on the environment it is being used in?
39. Are environmental requirements clearly defined?
40. Does the device and/or application remind users to be aware of their surroundings?
41. Is the language that is used in the device and/or application easy to understand?
42. Are sans serif font types used, as appropriate, throughout the device and/or application?
43. Is the contrast between the background and text sufficient enough that the text can be read easily under a range of normal lighting conditions?

44. If the text background is transparent, is the text visible across different backgrounds and under a range of normal lighting conditions?
45. Does the application and/or device use inclusive text and images?
46. Is the volume adjustable so the user can hear audio, even in noisy environments?
47. Are auditory features understandable?
48. Are captions available for auditory features as appropriate?

Heuristic 5: Integration of Physical & Virtual Worlds

49. Do the virtual elements help the user accomplish the required tasks in a meaningful way?
50. Is the visual appearance of the real-world environment sufficient to help the user accomplish required tasks?
51. Are physical (real-world) elements easily distinguishable from virtual elements?
52. Is it clear which virtual elements can be interacted with and which cannot?
53. Are virtual elements accurately placed on the real environment?
54. Is it clear how virtual objects relate to the real world environment?
55. Does the device's user interface and/or application avoid obstructing physical or virtual elements that are necessary for the users' goals?
56. Does the device's user interface and/or application avoid obstructing virtual navigation elements?

Heuristic 6: User Interaction

57. Does the user feel in control?
58. Are user interactions simple and easy to understand?

59. Does the device and/or application include multiple forms of interaction so users can choose based on ability, preference, & skill?
60. Are the forms of interaction direct when it is appropriate to use this form of interaction?
61. Does the device and/or application avoid interactions that force the user to make large or sudden movements?
62. Does the device and/or application accommodate for the user to complete other necessary real-world tasks?
63. Does object manipulation work well in all instances?
64. Do virtual elements adapt to the users' position appropriately?
65. Does the device and/or application avoid input overloading by assigning distinct functions to buttons or gestures?

Heuristic 7: Comfort

66. Can the user experience the device and/or application without pain, discomfort, nausea, disorientation, etc. DURING use?
67. Can the user experience the device and/or application without pain, discomfort, nausea, disorientation, etc. AFTER use?
68. Is the device's weight light enough to feel comfortable?
69. Does the device avoid overheating to the point that it is uncomfortable to use?
70. Does the device's accessories and cords avoid hindering work?
71. Are physical interactions with the application safe and comfortable?

72. Does the application avoid making the user walk backwards, pull their head back, or push their head downwards to see virtual elements?
73. Do interactions with the device and/or application avoid tiring the user?
74. Does the device and/or application avoid causing the user eye strain?
75. Are users reminded to take breaks?
76. Does the device easily adjust for a diverse set of users?
77. Does the device accommodate for personal protective equipment?

Heuristic 8: Feedback to the User

78. Does the device and/or application provide feedback on its status?
79. Does the device and its accessories provide feedback about battery levels and charging state?
80. Does the device and/or application provide feedback for user input?
81. Does the device and/or application respond quickly to user input?
82. Does the device and/or application provide the user feedback after automatic selections?
83. If an automatic selection occurs, does the device and/or application suggest what to do next?

Heuristic 9: Intuitiveness of Virtual Elements

84. Are virtual elements and icons self-explanatory (does their form communicate function)?
85. Are virtual elements and controls placed near objects they reference?

86. If a virtual element is related to an object that is in motion, is the virtual element tightly coupled with object in motion appropriately?
87. Are virtual elements that are outside of the field of view easy to find?
88. Are available user actions identifiable?
89. If voice commands are included, are text labels for voice commands given?

Heuristic 10: Collaboration

90. Are avatars representative of a diverse population?
91. If users are sharing the same virtual space, are virtual landmarks included to help orient users who may be in different physical spaces?
92. Is it easy to share virtual content to other users?
93. When collaborating with others, is it clear what content is and is not private?
94. Are avatars and virtual content rendered an adequate distance away from the user to preserve the user's personal space?
95. Does the device and/or application allow for the user to easily get to a private "safe place" in the event that other users are making them uncomfortable?
96. Is it clear what another user is referencing using non-verbal cues?
97. Is it clear which virtual elements can be interacted with and which cannot for each user?
98. Is content consistent across all users, as it is appropriate?

Heuristic 11: Privacy

99. Is it clear how user data is collected, stored, used, and protected?

100. Does the device and/or application allow for the user to control privacy-related content?

101. Does the device and/or application avoid those around the user to the utmost that the required tasks allow?

102. Is it clear to both users and bystanders when captures and/or recordings are being taken?

103. Is sharing virtual content to other users a mutual agreement?

104. Does the application allow users to preserve virtual elements from others users' changes?

Heuristic 12: Device Maintainability

105. Is the device sturdy enough to withstand multiple uses?

106. Does the device have a sturdy storage case?

107. Is it easy to clean the lenses, cameras, and other components on the device?

108. Are device parts fixable and replaceable as needed?

109. Does the device's battery life last long enough to perform necessary tasks of the application?

Appendix J

Definitions for the Revised 12 AR and MR Usability Heuristics

ID	**H1**
Priority	(2) Important
Name	Unboxing & Setting Up
Definition	Getting started with the AR/MR device/application should be easy to identify, complete, and a positive experience.
Explanation	A user's first experience is very important because it sets the tone and future experience with the device/application. If the set-up process is difficult, it could deter the user from any future use with the device.
Examples	When the user takes the device out of its packaging, it should be clear what each of the pieces do and how to assemble them (if necessary). Unboxing the device should be easy to do, the user should not have to struggle with taking the device out of its packaging. For a mobile experience, it should be clear how to interact with the AR/MR content (e.g., a call to action like a QR code or other identifying information).
Benefits	An easy unboxing experience can create a seamless and positive first experience. This can directly affect the user's first impressions with the device, setting a positive tone for future interactions.
Problems	Misunderstanding of this heuristic can cause the user to feel confused or frustrated and can lead to abandonment of use. If this unboxing and set-up process is difficult, there is potential that the device could be accidentally damaged.

ID	**H2**
Priority	(3) Critical
Name	Instructions
Definition	Help and documentation for the application and device should be easily accessible and easy to understand (including tutorials). Instructions and error messages should give users clear feedback.
Explanation	Help, documentation, instructions, and error messages are necessary after an issue has occurred. Users should be able to easily find, understand, and address all information that will help them resolve the issue quickly and easily.
Examples	An informative and easy to follow tutorial should be available when the user uses the device/application for the first time. This tutorial and other documentation should be available and easy to access during subsequent uses. Error messages should include actionable items to help the user recover from the errors such as, "move closer," "aim the device towards a flat surface," or "move to an area with more light"

Benefits	Even if a device/application is designed to be simple and intuitive, it is important that users have all documentation that can help them solve a problem if it occurs. This will increase the user's understanding of the device/application and help them complete all of the necessary tasks.
Problems	Misunderstanding this heuristic could lead users to make guesses about how to use the device/application. This could then lead to frustration and abandonment of use if they cannot locate information about questions they have.

ID	**H3**
Priority	(2) Important
Name	Organization & Simplification
Definition	The device/application should minimize cognitive overload by easing the user into the environment and avoiding unnecessary clutter.
Explanation	Reducing clutter, large amounts of text, and organizing information into an understandable manner helps users understand what they are supposed to interact with, how to complete tasks, and feel less overwhelmed by tasks.
Examples	There should be a clear layout with no distracting elements (e.g., large blocks of text, many different types of AR/MR elements obstructing the view of important physical objects that may be important for a task, etc.). The user should also be eased into the virtual environment (e.g., NOT telling a user to dodge projectiles as soon as they enter the game).
Benefits	When devices/applications have a minimalistic and aesthetically pleasing design, users are able to find information quickly and easily.
Problems	Misunderstanding this heuristic could potentially lead to a poor user experience with the application due to visual clutter and potential confusion of the important elements with irrelevant information.

ID	**H4**
Priority	(2) Important
Name	Consistency & Flexibility
Definition	The device/application should be consistent and follow design standards for text, audio, navigation, and other elements.
Explanation	Terminology and options should remain the same throughout the application to ensure that users do not get confused about their meaning. If there is a standard way to portray the information or interact with the device, these standards should be used to ensure a seamless experience.
Examples	Text should be large enough to read, controls should be based on known interactions (for mobile devices, already established gestures, such as "swipe" to scroll and "pinch" to zoom, should be used), malfunctions (such as lag, drift, and jitter) should be avoided, and menu items should look the same throughout the device/application.

Benefits	A consistent experience will allow the user to learn the device/application quickly and be less confused about their interactions with the device/application.
Problems	Misunderstanding this heuristic could lead to confusion and user errors. When features are inconsistent, users cannot develop clear expectations of how they should interact with the device/application.

ID	**H5**
Priority	(2) Important
Name	Integration of Physical & Virtual Worlds
Definition	It should be easy to identify virtual elements and which virtual elements are intractable. Virtual elements should not obstruct physical objects in the users' environment that are crucial for the completion of their goals.
Explanation	When there are multiple types of virtual elements in an AR/MR space (e.g., menu items, descriptions, holograms, etc.) it should be clear exactly which elements are able to be interacted with and which are only informative or are present for aesthetic. These virtual elements should flow well with the physical world. They should not obstruct physical objects that are safety risks (e.g., cords on the ground) or important for the task at hand.
Examples	Virtual elements should not block menu options (or other virtual navigational elements), potential risks in the physical environment (e.g., cords on the ground or low hanging materials), or physical objects that are important for the user's goals (e.g., machinery, an instruction pamphlet, etc.)
Benefits	A seamless integration of physical & virtual worlds can help the user learn how to interact with the virtual content and complete real-world tasks. It can also increase the user's immersion with the device.
Problems	Misunderstanding of this heuristic could cause the user to be confused and unsure about how to interact with the virtual elements. A risk to user safety may occur if the virtual elements obstruct physical hazards.

ID	**H6**
Priority	(3) Critical
Name	User Interaction
Definition	All interactions that the user has with the device/application should be simple, easy to understand, and easy to complete.
Explanation	User interactions should be easy to learn, and quick, safe, and comfortable to perform. Different ways of interaction should be available so the user can choose the method that works best for them. The user should feel confident about their control of the application and the virtual elements within it.

Examples	Allow for multiple ways to interact with an object (e.g., allow the user to "resize" an object through a voice command, one-handed gesture, two-handed gesture, or eye gaze selection).
Benefits	Quick and easy to learn user interactions can decrease the amount of cognitive workload and number of errors that a user experiences while also increasing efficiency and user satisfaction.
Problems	Misunderstanding of this heuristic can cause the device/application to become more difficult to learn and frustrating for the user. It also may result in the abandonment of use.

ID	**H7**
Priority	(3) Critical
Name	Comfort
Definition	The application and device should be designed to minimize user discomfort.
Explanation	Users should not experience any pain, discomfort, nausea, or disorientation as a result of using AR/MR. The device/application should minimize eye strain, remind the user to take breaks, and provide information about potential risks of discomfort to the user.
Examples	Include pop-ups that remind the user to take breaks, design the AR/MR device to be lightweight, allow for the use of protective equipment if it is necessary (such as hard hats, gloves, and/or eye protection), design for diversity (e.g., accommodate for different head sizes, face shapes, hair-styles, corrective lenses, etc.), and avoid unsafe or uncomfortable interactions (e.g., large swiping gesture that involves the movement of the entire arm while in a crowded manufacturing environment, or interactions that force the user to look upward for long periods of time).
Benefits	Creating a device/application for comfort can include more users to participate in the use of the device/application and make it a more enjoyable experience.
Problems	Misunderstanding of this heuristic can cause the user to feel uncomfortable and can potentially harm the user.

ID	**H8**
Priority	(2) Important
Name	Feedback to the User
Definition	The application and device should provide adequate feedback to the user to explain what is currently going on.
Explanation	When a user or device/application is completing an action, the device/application should provide some sort of feedback (visual, auditory, etc.) to the user that indicates that the action was successful. Likewise, if there is an issue while completing the action, feedback should be given to the user explaining why the action could not be completed.

Examples	Include a visual representation while scanning the environment and during loading screens. If an automatic selection occurs (e.g., when a user fails to select an option within a reasonable amount of time) provide the user with feedback about the option that was chosen and what to do next.
Benefits	Providing feedback to the user can help the user understand what the device/application is doing and how to proceed. This reduces the amount of confusion that they may have when interacting with the device/application and sets expectations about how it should perform.
Problems	Misunderstanding of this heuristic can cause confusion and frustration for the user. It may also result in errors that users may not understand how to recover from.

ID	**H9**
Priority	(2) Important
Name	Intuitiveness of Virtual Elements
Definition	The application should be designed in a way that promotes the use of recognition rather than recall to minimize the user's memory load.
Explanation	Providing too much information or novel concepts that need to be remembered to either use the device/application or complete a later task is cognitively taxing for the user. The user should not have to remember information from one part of the device/application to another. Encouraging the use of recognition by providing guidance and recognizable elements rather than forcing the user to recall all previous information will ease cognitive load.
Examples	If virtual elements are off-screen, include arrows (or other form of identification) to grab the user's attention to it. Virtual tags (such as those that label machine parts or other physical or virtual objects) should accurately follow the parts as the user moves and manipulates them.
Benefits	Minimizing workload can reduce user's mental fatigue, help users complete tasks effectively and efficiently, and make it easier for users to learn the device/application quickly.
Problems	Misunderstanding of this heuristic can lead to higher cognitive workload and can result in user errors and frustration.

ID	**H10**
Priority	(2) Important
Name	Collaboration
Definition	When sharing an AR/MR space with others, it should be easy to understand what actions are available, what is private vs. public, and communication between users should be seamless.
Explanation	Sometimes AR/MR applications can include multiple users, either with the same abilities or different abilities than each other. It should be clear how

	communication and teamwork can occur in these spaces (both the abilities and the limitations of available interactions). It is also important to allow users to have a private space in case they do not want to share all content. Finally, avatars should represent users accurately (be representative of a diverse population as well as the user's current actions).
Examples	Avatars that represent the users; communication includes non-verbal as well as verbal cues to make the experience more natural; the ability to "lock" virtual elements (so users other than the creator of the content cannot delete them).
Benefits	Including avatars that represent a diverse set of users can make the experience enjoyable and more relatable across users. Making interactions and expectations clear between multiple users sharing an experience can increase communication, task success, and enjoyment between users.
Problems	Misunderstanding this heuristic could cause users to become frustrated or mistrustful of a device or application if it is no clear what is shared to others, how to communicate with others, and/or losing autonomy to others' control of the virtual content.

ID	**H11**
Priority	(2) Important
Name	Privacy
Definition	User and bystander privacy should be addressed in the design of the device/application. It should be clear how data is being used and protected.
Explanation	The user's privacy and the privacy of those around them should be considered when designing AR/MR applications/devices. This includes personal information (e.g., location, biometric data, etc.) and images of themselves (e.g., a user recording someone around them). This is important for ethical considerations, legal requirements that users have to adhere to for their AR/MR use cases, and to instill trust in the system.
Examples	If a user's biometric data or location is being collected, they should be aware of how that is being collected and used (e.g., a pop-up that provides additional information before using the application for the first time). If the device can record bystanders, it should be clear to bystanders when that is happening (e.g., a light appears on the device).
Benefits	Knowing how data is being collected, used, and protected can instill user trust in the system.
Problems	If user and bystander privacy is not being protected, it could lead to abandonment of use either due to the user not trusting the system or because the user may have to follow strict guidelines and is unable to use the system if privacy is not a priority in the application/device. This is often the case for government use cases.

ID	**H12**

Priority	(1) Useful
Name	Device Maintainability
Definition	The device should be designed in a way that makes it easy to maintain. This includes reusability, storage, cleaning, and the ability to fix/replace parts.
Explanation	The device should be able to be used multiple times. During this reuse, it is possible that the device may get dirty, and parts may break. It is important that the process to clean and fix/replace parts is easy (and occurs infrequently) so the user can trust the reliability of the device.
Examples	If a part (such as a strap) breaks, it should be easy to purchase and install a new one.
Benefits	It is very likely that parts will need to be cleaned or replaced after many instances of reuse. User satisfaction can increase if you give control to users by giving them the knowledge and access to clean and replace parts themselves.
Problems	Misunderstanding of this heuristic can cause potential damage to the device and user frustration.