Effectiveness and User Experience of Augmented and Mixed Reality for Procedural Task Training

Emily Rickel
rickele@my.erau.edu

Follow this and additional works at: https://commons.erau.edu/edt

Part of the Applied Behavior Analysis Commons, and the Human Factors Psychology Commons

Scholarly Commons Citation
https://commons.erau.edu/edt/725

This Dissertation - Open Access is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in Doctoral Dissertations and Master's Theses by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.
Effectiveness and User Experience of Augmented and Mixed Reality for Procedural Task Training

By

EMILY A. RICKEL

Bachelor of Science in Human Factors Psychology, Embry-Riddle Aeronautical University, 2019

Master of Science in Human Factors, Embry-Riddle Aeronautical University, 2020

A dissertation submitted 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
Spring 2023
Acknowledgements

I would like to express my deepest gratitude to my advisor and dissertation chair, Dr. Barbara Chaparro. Thank you for your endless encouragement and dedication in promoting my growth as a researcher. I am thankful that you were always approachable and that you readily squashed any obstacle that stood between me and the finish line. I am also immensely grateful for the support of my other dissertation committee members, Ms. Beth Atkinson, Dr. Beth Blickensderfer, and Dr. Joseph Keebler. I appreciate your willingness to share your valuable expertise and guidance not only throughout the process of completing my dissertation, but also throughout my years as a student at ERAU.

Many thanks to Darrin Knaggs, Reid Santiago, and Zachary Phillips for eagerly and diligently assessing hundreds of origami models. Thank you to my ERAU peers and NAWCTSD mentors for sharing their wisdom and support, and for helping me celebrate all the milestones along the way. A big thank you to those who offered their assistance as pilot participants and avid sounding boards, including Shivani and John, and my academic siblings, Carmen and Jess.

Finally, thank you to my family and friends who offered their continuous love and support from across the state, country, and continent. I appreciate and love all of you. Special thanks to my parents, who have been my biggest cheerleaders for as long as I can remember. And thank you to Jack, who has steadfastly been by my side throughout this academic journey. You’re the best everything buddy.
Abstract

Use of augmented reality (AR) and mixed reality (MR) technologies for training is increasing, due in part to opportunities for increased immersion, safer training, and reduced costs. However, AR/MR training effectiveness and user experience, particularly for head-mounted displays (HMDs), is not well understood. The purpose of this study is to investigate user perceptions and retention of AR/MR training delivered through a HMD for a procedural task. This two-part study utilized a within-subjects experimental design with 30 participants to determine how instruction method (paper vs. AR vs. MR) and time of procedure recall (immediate vs. post-test vs. retention) influenced completion time, perceived task difficulty, perceived confidence in successfully completing the task, workload, user experience, and trainee reactions. Results indicate differences between instruction methods for user experience and preference, with significantly higher user experience ratings for MR and lower preference rankings for AR. Findings also show decreased performance, increased perceived task difficulty, and decreased confidence as time since training increased, with no significant differences in these measures between instruction methods. Completion times and workload were also found to be comparable between instruction methods. This work provides insight into objective and subjective differences between paper-, AR-, and MR-based training experiences, which can be used to determine which type of training is best suited for a particular use case. Recommendations for appropriately matching training modalities and scenarios, as well as for how to successfully design AR/MR training experiences, are discussed.

Keywords: augmented reality, mixed reality, training effectiveness, user experience, procedural task
# Table of Contents

Literature Review ................................................................. 1

Extended Reality ........................................................................ 1
  Definitions of Extended, Virtual, Augmented, and Mixed Reality .......... 1
  The Evolution of Extended Reality Technologies ............................... 3

Training...................................................................................... 6
  Maximizing Training Effectiveness ............................................... 6
  Expanding on Training Evaluation ............................................. 7

Training Procedural Tasks ....................................................... 10
  Origami as a Procedural Training Task ....................................... 13

Extended Reality Training ....................................................... 14
  Defining Extended Reality Training .......................................... 15
  Extended Reality Training Opportunities ................................... 15
  Learning Theories Supporting Extended Reality Training ............... 17
  Extended Reality Training Challenges ...................................... 20
  Extended Reality Training Effectiveness .................................... 23

Current Literature Gaps .......................................................... 29

Purpose .................................................................................... 30

Research Design ......................................................................... 31

Hypotheses .............................................................................. 32
  Performance ........................................................................... 32
  Time to Completion .................................................................. 33
  Perceived Difficulty .................................................................. 33
  Perceived Confidence ................................................................ 34
  Workload ................................................................................ 34
  User Experience ....................................................................... 35
  Trainee Reactions ..................................................................... 36
  Exploratory Hypotheses .......................................................... 36

Method ...................................................................................... 37

Participants ............................................................................... 37

Materials ................................................................................... 37
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptions on Whether Instruction Method Matters</td>
<td>81</td>
</tr>
<tr>
<td>Results Summary</td>
<td>82</td>
</tr>
<tr>
<td>Discussion</td>
<td>85</td>
</tr>
<tr>
<td>Theoretical Implications</td>
<td>86</td>
</tr>
<tr>
<td>Practical Implications</td>
<td>88</td>
</tr>
<tr>
<td>Limitations</td>
<td>89</td>
</tr>
<tr>
<td>Future Research</td>
<td>91</td>
</tr>
<tr>
<td>References</td>
<td>92</td>
</tr>
<tr>
<td>Appendices</td>
<td>108</td>
</tr>
<tr>
<td>Appendix A</td>
<td>108</td>
</tr>
<tr>
<td>Appendix B</td>
<td>109</td>
</tr>
<tr>
<td>Appendix C</td>
<td>112</td>
</tr>
<tr>
<td>Appendix D</td>
<td>115</td>
</tr>
<tr>
<td>Appendix E</td>
<td>117</td>
</tr>
<tr>
<td>Appendix F</td>
<td>118</td>
</tr>
<tr>
<td>Appendix G</td>
<td>119</td>
</tr>
<tr>
<td>Appendix H</td>
<td>120</td>
</tr>
<tr>
<td>Appendix I</td>
<td>121</td>
</tr>
<tr>
<td>Appendix J</td>
<td>122</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: AR, MR, and VR Placed Along a Reality-Virtuality Continuum ................................................................. 1
Figure 2: Examples of Current XR Use Cases .................................................................................................................. 5
Figure 3: The Kirkpatrick Model of Training Evaluation .............................................................................................. 9
Figure 4: Model of Processing Procedural Instructions .................................................................................................. 12
Figure 5: Origami Diagram Example ............................................................................................................................. 13
Figure 6: The Experiential Learning Cycle .................................................................................................................... 19
Figure 7: Appearance of Paper Instructions .................................................................................................................. 39
Figure 8: Appearance of AR and MR Instructions ........................................................................................................ 40
Figure 9: Microsoft HoloLens 2 ...................................................................................................................................... 41
Figure 10: Microsoft Dynamics 365 Guides ................................................................................................................... 43
Figure 11: Study Station ................................................................................................................................................ 45
Figure 12: Procedure Sequence ...................................................................................................................................... 52
Figure 13: Number of Hours of Reported XR Headset Use Prior to Study ................................................................. 53
Figure 14: Self-Reported Origami Familiarity ................................................................................................................ 55
Figure 15: Self-Reported Origami Proficiency ............................................................................................................... 55
Figure 16: Comparison of Performance Between Instruction Methods Over Time ................................................. 59
Figure 17: Comparison of Completion Time Between Instruction Methods Over Time .............................................. 60
Figure 18: Comparison of Perceived Instruction Method Difficulty ............................................................................. 61
Figure 19: Comparison of Perceived Task Difficulty Between Instruction Methods Over Time ............................. 62
Figure 20: Comparison of Perceived Difficulty Between Origami Models ............................................................... 63
Figure 21: Comparison of Perceived Confidence Between Instruction Methods Over Time .................................... 65
Figure 22: Comparison of Perceived Confidence Between Origami Models ............................................................... 66
Figure 23: Comparison of Workload Ratings Between Instruction Methods .......................................................... 67
Figure 24: Comparison of User Experience Ratings Between Instruction Methods .................................................... 68
Figure 25: Comparison of Trainee Reaction Ratings Between Instruction Methods .................................................. 70
Figure 26: Comparison of Preference Between Instruction Methods ........................................................................... 76
List of Tables

Table 1: Summary of Recent XR Training and Education Review Papers .................................................. 25
Table 2: Study Conditions ......................................................................................................................... 31
Table 3: Participants’ Self-Reported Video Game Experience ...................................................................... 54
Table 4: Number of Days Between Participants’ First and Second Study Session ........................................ 56
Table 5: Measures for Each Study Condition ............................................................................................ 57
Table 6: Trainee Reactions Questionnaire Items ....................................................................................... 69
Table 7: Participant Likes for Each Instruction Method ............................................................................... 71
Table 8: Participant Dislikes for Each Instruction Method ......................................................................... 72
Table 9: Correlation Matrix for Select Demographic Variables and Performance ...................................... 75
Table 10: Visualization Strategies Used by Participants to Recall the Retention Models .......................... 80
Table 11: Other Strategies Used by Participants to Recall the Retention Models ....................................... 81
Table 12: Summary of Hypotheses and Outcomes of Performed Analyses .............................................. 83
Table 13: Summary of Study Results ....................................................................................................... 85
Extended Reality

Definitions of Extended, Virtual, Augmented, and Mixed Reality

**Extended Reality.** Extended reality (XR) is a collective term referring to technologies that combine varying degrees of real and virtual elements to generate immersive human-computer experiences (Marr, 2021; Stanney et al., 2021). It is difficult to pinpoint standardized definitions of XR technologies because they are rapidly evolving. However, it is commonly accepted that XR includes several different experiences, including virtual reality (VR), mixed reality (MR), and augmented reality (AR), under its umbrella (Rauschnabel et al., 2022). These experiences can be compared by placing them on a continuum ranging from the real environment to the virtual environment. Figure 1 displays where VR, MR, and AR fall along such a continuum, as well as a few distinguishing traits of each technology.

**Figure 1**

*AR, MR, and VR Placed Along a Reality-Virtuality Continuum*

**Augmented Reality (AR)**
- Maintain visibility of physical environment
- Able to interact with virtual elements

**Mixed Reality (MR)**
- Maintain visibility of physical environment
- Able to interact with and manipulate virtual elements

**Virtual Reality (VR)**
- Full immersion in digital environment
- Able to interact with and manipulate virtual elements

*Note. Adapted from Milgram & Kishino (1994) and Derby et al. (2020).*
**Virtual Reality.** Sherman & Craig (2003) define VR as “a medium composed of interactive computer simulations that sense the participant’s position and actions and replace or augment the feedback to one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world),” (p. 16). According to Sherman & Craig (2003), the four key elements of VR experiences are a virtual world, immersion, interactivity, and responding to the user’s input. VR involves full immersion in the digital environment (Marr, 2021). This immersion is typically generated using a head-mounted display (HMD; e.g., Oculus Quest, HTC Vive) or virtual environment room (e.g., CAVE) that blocks out visual stimuli from the real world. In VR, users can interact with and manipulate virtual elements within their digital environment through the use of several user input modalities, including controllers, gloves, eye-tracking, or voice commands (Sherman & Craig, 2003).

**Mixed Reality.** Similar to VR, MR also allows users to manipulate virtual elements. However, unlike VR, virtual elements in MR can be manipulated while the user maintains visibility of their physical surroundings (Marr, 2021). Additionally, MR further blends the real and virtual worlds by embedding responsive digital elements within the user’s physical space (Brigham, 2017), allowing digital content to acknowledge and interact with the real world (Stanney et al., 2021). For example, MR HMDs like the Microsoft HoloLens 2 work by scanning, mapping, and superimposing the user’s surroundings with virtual objects. These virtual objects can be anchored to physical landmarks, and obscured by physical objects (Microsoft, 2022a).

**Augmented Reality.** AR is similar to MR in that users maintain visual awareness of the real world. However, digital content in AR is simply overlaid onto the user’s view of the real world and is not responsive to physical elements. As a result, AR users will not experience depth and perspective of virtual objects, as MR users would (Brigham, 2017). Thus, when mapping AR, VR, and MR to the reality-virtuality continuum (see Figure 1), AR is more representative of the real environment, while VR is more representative of the virtual environment, and MR falls in the middle.
The Evolution of Extended Reality Technologies

The origins of XR technologies can be traced back to the 1830s, when Charles Wheatstone developed the first stereoscope (i.e., a device that displays a pair of slightly different images, one to each eye, to create the illusion of a three-dimensional image; Marr, 2021; Pope, 2018). Stereoscopic displays are used in many of today’s XR systems to promote the feeling of immersion by creating a sense of depth in digital content (Marr, 2021; Stanney et al., 2021). Between the 1830s and the 1960s, science fiction writer Stanley Weinbaum penned a story describing a character who explores a fictional world through a pair of goggles, an experience that mirrors many of today’s VR systems. Additionally, cinematographer Morton Heilig developed an immersive movie booth with a stereoscopic screen, stereo speakers, scent releasers, and vibrating seats. Heilig later went on to patent the first VR HMD in 1960. Dubbed the Telesphere Mask, the primary purpose of Heilig’s HMD was to show movies and did not incorporate motion-tracking capabilities. Within a year, engineers from Philco Corporation released the first motion-tracking VR HMD. XR innovations in the 1960s continued under the work of computer scientist Ivan Sutherland, who created what is widely considered to be the blueprint for modern VR systems. Sutherland also created the first AR headset in 1968 (Marr, 2021; Pope, 2018).

In the 1980s and 1990s, the terms “virtual reality” and “augmented reality” were coined by Jaron Lanier and Thomas Caudell, respectively. Paul Milgram and Fumio Kishino (1994) also introduced the term “mixed reality” during this time. These decades also saw the development of VR gloves that provided users the ability to control virtual content with hand gesture input. Additionally, 3D video game consoles and VR headsets emerged for consumer use (Marr, 2021; Pope, 2018). However, these first attempts at tapping into the consumer market failed due to cost and technical difficulties, such as bulky hardware, restricted processing power, and low resolution and frame rates (Hillmann, 2021; Pope, 2018). The 1990s was also an important time for AR, when sports games started to overlay graphics on top of the game’s live camera feed (Marr, 2021; Pope, 2018).
While the 2000s were a relatively quiet time in XR history (Marr, 2021), the 2010s are considered a defining decade for modern XR capabilities (Hillmann, 2021). During this time, big tech companies began to relaunch VR headsets geared towards consumers (e.g., Oculus Rift, later acquired by Facebook; Google Cardboard, a low-cost device that allows users to turn their own smartphones into a VR viewer). In 2016, the release of AR mobile gaming app Pokémon Go and its subsequent skyrocketing popularity marked the first mainstream success for AR. In the same year, Microsoft developed their first MR headset, the HoloLens. By the end of the 2010s, hundreds of companies across a variety of industries (e.g., retail, manufacturing, tourism, journalism, marketing, social media) were developing XR experiences for consumer and enterprise use cases (Hillmann, 2021; Marr, 2021). This explosive growth can be attributed to technological advancements that promote mass adoption of XR systems, including higher frame rates, extended battery life, increased mobile bandwidth, decreased data latency, lower device cost, and sleeker hardware designs (Cook et al., 2017; Stanney et al., 2021).

Today, in the early 2020s, XR systems are rapidly evolving to become faster, more responsive, and more portable due to developments in artificial intelligence, cloud computing and storage systems, and network speeds (Marr, 2021; Stanney et al., 2021). The COVID-19 pandemic was another turning point for XR. Unable to interact in the physical world, many individuals and organizations adopted XR solutions for virtual meet-ups (Hillmann, 2021; Koumaditis et al., 2021). As shown in Figure 2, XR applications are currently being utilized in a wide range of domains, including retail (Kumar, 2022; Rickel & Roa, 2020), social media and marketing (Novakova & ŠTarchoň, 2021), gaming and entertainment (Pu et al., 2022), tourism (Weber-Sabil & Han, 2021), medicine (Barteit et al., 2021; Venkatesan et al., 2021), and training (Kaplan et al., 2021). Future market predictions estimate that worldwide spending on XR technologies will increase from $12 billion in 2020 to $72.8 billion in 2024 (International Data Corporation, 2020).
Figure 2

Examples of Current XR Use Cases

Mobile AR for Retail

MR HMD for Medicine

VR HMD for Flight Training

Training

The term “training” can be defined as a systematic effort to transmit knowledge, skills, and attitudes (KSAs) with the goal of improving performance (Bisbey et al., 2021). Training involves administering an intervention to produce sustainable changes in individuals’ behavior and cognition (Salas et al., 2012). Effective training consists of providing trainees with instruction, demonstration, practice, and feedback about their performance (Salas & Cannon-Bowers, 2001). Training helps organizations achieve goals (e.g., improve performance, reduce errors, promote safety) and maintain a competitive advantage by facilitating workforce learning and development (Salas et al., 2012). In 2021, U.S.-based corporations and educational institutions spent approximately $92.3 billion on training expenditures (Freifeld, 2021), a significant investment indicating that organizations recognize the importance of training their workforce.

Maximizing Training Effectiveness

Training effectiveness refers to individual, training, and organizational traits that influence the likelihood of achieving successful training outcomes (Alvarez et al., 2004). Salas et al. (2012) and Bisbey et al. (2021) provide recommendations for maximizing training effectiveness by outlining steps that should be taken before, during, and after training:

Before Training. Prior to developing a training program, it is recommended to conduct a training needs analysis to understand what needs to be trained and who needs the training. This analysis should inform learning objectives, training delivery and evaluation methods, as well as individual and organizational factors that may impact training effectiveness.

During Training. Several aspects of how training is delivered can impact training effectiveness, including trainee characteristics and instructional strategies. Trainee characteristics such as self-efficacy (i.e., what trainees think about their own abilities) and motivation to learn (i.e., the interest and effort that trainees impart on training) influence how much trainees learn during training. Because higher self-
efficacy has shown to promote learning, training should be designed to facilitate and reinforce trainees’ beliefs in their abilities. Also positively correlated with learning, motivation to learn should be encouraged throughout training. Methods for encouraging motivation to learn include solidifying the connection between training content and trainees’ job demands, as well as providing organizational and supervisory support. Furthermore, increased learning typically occurs when the following tools and methods (i.e., instructional strategies) are used to deliver training: 1) engaging and knowledgeable instructors, 2) demonstrations of the desired behaviors and cognitions, 3) opportunities for practice, 4) constructive and timely feedback, and 5) selecting delivery systems (i.e., classroom lecture, computer-based, simulation) that complement the training objectives and content.

**After Training.** Post-training activities can influence training effectiveness. After training, transfer of training (i.e., the extent to which trained KSAs are applied to the trainee’s job) can be facilitated by ensuring trainees can reinforce learned KSAs while on the job, clarifying the importance of applying training to the job, and providing trainees with aids to promote recall of training content. Training should also be evaluated to determine whether it was effective. Such an evaluation helps organizations know whether to continue conducting the training, or whether the training needs to be modified or discontinued.

**Expanding on Training Evaluation**

Training evaluation involves the collection of data to determine the success of a training program. Training evaluations aim to measure whether learning objectives were achieved by trainees and whether learning those objectives resulted in on-the-job performance improvements (Kraiger et al., 1993). Alvarez et al. (2004) distinguishes the terms “training evaluation” and “training effectiveness”, such that a “training evaluation” identifies the extent to which the training was successful in meeting its intended goals, while “training effectiveness” findings provide an explanation as to why the training was successful or unsuccessful and how the training can be improved. Training effectiveness studies involve
manipulating variables that may facilitate or hinder the likelihood of achieving successful training outcomes. For example, a training effectiveness study could consist of developing and comparing multiple training programs with differing characteristics (Alvarez et al., 2004).

Common outcomes measured during training evaluations include cognitive, skill-based, and affective learning outcomes (Kraiger et al., 1993), as well as organizational payoffs (Alvarez et al., 2004). Cognitive learning outcomes gauge knowledge acquisition by evaluating verbal knowledge, knowledge organization (i.e., mental models), and cognitive strategies (i.e., methods for accessing or applying knowledge more quickly). Cognitive outcomes are commonly measured using recognition and recall tests. Skill-based learning outcomes assess the evolution of motor skills. Observation (i.e., tracking frequency of desired or undesired behaviors, step sequence, errors, or time to completion as trainees demonstrate completing a task) and interviews (i.e., asking trainees to describe how they would complete a task) are popular methods for measuring skill-based outcomes. Affective learning outcomes appraise changes to a trainee’s attitude, motivation, self-efficacy, and goal setting as a result of completing the training. Affective outcomes are typically measured through self-report methods (Kraiger et al., 1993). Organizational outcomes vary based on the organization’s goals, but may include productivity, quality, safety, sales, customer satisfaction, and employee turnover (Alvarez et al., 2004; Bisbey et al., 2021; Kirkpatrick & Kirkpatrick, 2016; Topno, 2012).

Originally proposed in the 1950s, the Kirkpatrick Model is today’s most widely-used framework for structuring training evaluations (Bisbey et al., 2021; Kirkpatrick & Kirkpatrick, 2016; Tamkin et al., 2002; Tripathi & Artibansal, 2017). The model organizes training evaluations into four levels: 1) reaction, 2) learning, 3) behavior, and 4) results. Level 1, Reaction, refers to what trainees think about the training. Satisfaction with training is often measured immediately after training using self-report methods such as surveys, questionnaires, or interviews. It is recommended to create self-report items structured around the organization’s needs and training program’s goals, as a standardized evaluation
form that can be applied to all circumstances does not exist (Kirkpatrick & Kirkpatrick, 2016). It is recommended to design self-report items using closed-ended questions (e.g., rating scales, multiple choice questions) in order to be able to quantify results, as well as open-ended questions in order to acquire additional comments that may not be collected by the closed-ended questions (Kirkpatrick, 1967). Level 2, Learning, captures the extent to which trainees achieved the intended KSAs. Evaluating knowledge and skills through post-training tests and demonstrations are some of the most popular methods for Level 2 evaluations. Level 3, Behavior, measures the extent to which trainees apply what they learned to their job. Creating and monitoring an action plan for how to incorporate things learned during training, as well as observations, are common Level 3 evaluation methods. Level 4, Results, assesses the impact of the training program on the organization. Tracking changes in desired business metrics (e.g., productivity, safety, customer satisfaction) is often the evaluation method of choice for Level 4 (Kirkpatrick & Kirkpatrick, 2016). As shown in Figure 3, the four levels represent a sequential hierarchy of evaluation steps, where it is assumed that advancing to the next step is more difficult and requires more resources, but doing so also provides more valuable information (Reio et al., 2017).

Figure 3

*The Kirkpatrick Model of Training Evaluation*

Strengths of the Kirkpatrick Model that may explain its popularity include its ability to simplify the complex process of training evaluation (Bates, 2004). This simplicity translates to practicality by providing a framework that is easy to implement in applied, real-world training situations. Additionally,
the model is broad and flexible enough to apply to a variety of organizations and training programs (Kirkpatrick & Kirkpatrick, 2016). However, there are several critiques of the model. Kraiger et al. (1993) reason that the Kirkpatrick Model is too simple, presenting a unidimensional definition of the multifaceted concept of learning. Alliger & Janak (1989) also argue that it is problematic to assume the four levels build upon one another to provide more valuable information (e.g., that measures of Behavior provide more valuable information than measures of Learning), as this can lead to the belief that Level 4 is the most valuable measure. As a result, those in charge of overseeing training evaluations may opt to skip the lower levels and focus only on the higher levels (Reio et al., 2017). Additionally, it may be improper to assume causality between levels, or to perceive they are positively correlated. In practice, training evaluation instruments may attempt to measure across multiple levels at the same time. This lack of temporal distinction in level assessment provides evidence against the causal relationship between levels. Furthermore, levels may not be positively correlated. For example, trainees may react positively to training without learning anything, or vice versa (Alliger & Janak, 1989). Additional research is needed to mitigate these limitations.

Training Procedural Tasks

The term “procedural knowledge” can be defined as the ability to solve problems through the execution of action sequences (Rittle-Johnson et al., 2001). In other words, procedural knowledge refers to knowing how to do something (Krathwohl, 2002). Performing a task that involves procedural knowledge requires the recall of learned skills and behaviors, rather than facts (Sternberg & Sternberg, 2012). Additionally, completing a procedural task likely necessitates knowledge of how many steps are needed to perform the task, as well as what must be done at each step and the order in which the steps must be executed (Hochmitz & Yuviler-Gavish, 2011). Procedural knowledge is best acquired through practice. Practice can also increase the speed and accuracy at which procedural tasks are completed (Sternberg & Sternberg, 2012).
Procedural task training can be facilitated by providing trainees with training aids, such as instructions (i.e., documentation that explains the steps required to complete a task). Instructions describe system states and the actions necessary for moving from one system state to another (Eiriksdottir & Catrambone, 2011). Procedural task instructions are often organized as a linear sequence of steps leading to a single overarching goal (Konoske, 1985). Instructions with diagrams or pictures promote procedural instruction processing (Ganier et al., 2000) and may reduce cognitive load because it is easier to match pictures to real-world objects compared to using text-only instructions (Ganier, 2004).

Ganier et al. (2000) proposed a model of processing procedural instructions (see Figure 4). This model focuses on how users read, understand, and apply instructions for the first time, as opposed to the sub-processes involved in acquiring and storing procedural tasks in long-term memory. The model indicates that processing instructions begins with setting a goal and determining whether this goal matches the objective(s) outlined in the instructional document. As users inspect the instructions, they form a mental model of the task and its steps. Forming this mental model involves integrating information from the instructions and equipment, as well as retrieving prior knowledge from the user’s long-term memory storage. The accuracy and speed at which the user forms the mental model depends on the number of inferences the user has to make based on their prior knowledge, the equipment and its affordances, and how the instructions are presented. Next, users will construct an action plan based on their mental model that specifies the sequence of actions required to complete their goal. Then, users will monitor and regulate their action plan. Monitoring an action plan involves determining whether the state of the equipment matches the user’s mental model of the instructions and anticipated equipment state. The comparison of the user’s mental model and the equipment state is repeated until the initial goal is achieved.

The Ganier et al. (2000) Model of Processing Procedural Instructions illustrates the amount of attentional switching required for trainees to complete procedural tasks. Trainees must attend to the instructions, their mental model of the instructions, their action plan progress, and their prior knowledge stored in long-term memory, as well as to the equipment state. Attending to and processing such a large amount of information can increase trainees' mental workload. Strategies for minimizing mental workload while following procedural instructions include designing the instructions to be easily navigable (e.g., use of clear and prominent headings, including an index based on potential trainee goals for more complex procedures), as well as presenting instructions in both text and picture format to
facilitate mental model formation and reduce attentional switching between the instructions and equipment.

**Origami as a Procedural Training Task**

Derived from two Japanese words, “ori” (i.e., “folded”) and “kami” (i.e., “paper”), origami is the art of paper folding (Georgia Technical Institute of Technology, n.d.a). Instructions for origami are commonly presented as a series of diagrams that portray folding sequences. These diagrams are typically designed to depict as much information as possible in as few steps as possible (Robinson, 2016). Origami diagrams often incorporate symbology consisting of lines and arrows. As shown in Figure 5, solid lines either represent the paper’s edges or a crease where the paper has been previously folded. Dashed lines signify where the paper is to be folded. Arrows indicate which direction to fold the paper (Georgia Technical Institute of Technology, n.d.b).

**Figure 5**

*Origami Diagram Example*

Prior research has shown that origami is a suitable task for studying the acquisition of procedural skills (Novick & Morse, 2000; Tenbrink & Taylor, 2015; Wong et al., 2009; Zhao et al., 2020). Completing an origami model requires controlled, accurate movements. Performing steps incorrectly or out of order can prevent one from achieving a neat, attractive result, and can even obstruct the ability to finish a model (Robinson, 2016; Zhao et al., 2020). These task traits can facilitate a more straightforward and sensitive assessment based on accuracy throughout the procedure, rather than just a binary success/fail rating after the procedure. Another motive for using origami to study procedural training is that it comprises common features of sequential skills and procedural processing. Origami requires individuals to organize their behavior according to goals and action plans (Zhao et al., 2020), similar to processes outlined in the Ganier et al. (2000) Model of Processing Procedural Instructions. Therefore, origami has the potential to be generalized to other sequential motor tasks. Furthermore, while most people have at least some familiarity with origami, their paper folding experience is likely limited. As a result, it is easy to find novice participants for training studies (Novick & Morse, 2000; Zhao et al., 2020). Additionally, selecting a more general task like origami can minimize bias towards populations with expertise in specific tasks (Volmer et al., 2018). Moreover, most people take pride in successfully constructing an object and, thus, would likely be motivated to complete an origami model correctly and conscientiously (Novick & Morse, 2000). Finally, origami requires materials (i.e., paper) that are easily accessible and cost-effective, making it an economical task for research projects (Zhao et al., 2020).

Extended Reality Training

According to Training Magazine’s 2021 Training Industry Report, around 5% of surveyed organizations currently utilize XR training delivery methods. Considered a newer training delivery method, it was also reported that larger companies (i.e., 10,000 or more employees) were more likely to invest in XR technologies. Focusing only on larger companies, around 16% reported using VR and 13%
are using AR (Freifeld, 2021). Over the next few years, training is anticipated to be one of the most popular commercial use cases for XR systems. Worldwide investments in XR training solutions are forecasted to total over $4 billion in 2024 (International Data Corporation, 2020). These remarkable investments suggest that organizations are rapidly adopting XR training solutions, a process that also requires a considerable amount of time. But, are these investments warranted? The following sections will define XR training, explore XR training opportunities, supporting theories, and challenges, as well as summarize prior literature regarding whether XR systems are effective training delivery methods.

**Defining Extended Reality Training**

As previously mentioned, XR is the umbrella term for technologies such as AR, MR, and VR. Due to the wide-ranging capabilities and rapid evolution of XR technologies, it can be difficult to provide a flexible, enduring definition of XR training. Palmas & Klinker (2020) attempt to define XR training by stating it is a purposely designed immersive learning experience that leverages technologies that engage and support trainees as they acquire the knowledge and skills needed to impact outcomes aligned with organizational goals. The authors elaborate that “purposely designed” refers to structuring content around specific learning outcomes and intentionally using principles of effective instructional design to support trainees’ learning processes. Additionally, an “immersive learning experience” refers to one where trainees can actively interact with the training content within environments that may range from completely physical to completely virtual (Palmas & Klinker, 2020).

**Extended Reality Training Opportunities**

There are several advantages to XR-based training, including reduced costs, safer training, enhanced flexibility, and opportunities for feedback, practice, and immersion.

**Reduced Costs.** While initial costs for XR training solutions may be more expensive (Farra et al., 2019; Thompson et al., 2009), long-term costs tend to be less than live exercises (Farra et al., 2019; Gonzalez-Franco et al., 2017; Kaplan et al., 2021; Thompson et al., 2009). Simulation-based training can
also reduce the number of hours in which the operational system is in use. This reduction in use lessens wear and tear on the operational system, decreasing maintenance and repair costs (Thompson et al., 2009). Additionally, XR technologies can simulate physical objects and operations that may be too expensive to utilize in training, enabling the interaction of content and procedures that were previously not available due to budget constraints (Minna et al., 2021).

**Safer Training.** XR makes it possible to train for inaccessible or unreachable scenarios, and in locations that are difficult to reproduce in the real world (Kaplan et al., 2021; Minna et al., 2021; Sowndararajan et al., 2008). Simulators can expose trainees to unsafe conditions that would be too dangerous to perform in a real-world setting to facilitate the development of technical and decision-making skills required for desired performance in hazardous situations (Martin et al., 2014; Thompson et al., 2009). Additionally, instructors can demonstrate or encourage trainees to perform procedures that may not be desired or permissible in operational environments in order to safely portray the consequences of completing such a procedure (Thompson et al., 2009).

**Enhanced Flexibility.** XR simulations are flexible and can be rapidly updated to reflect new information as it becomes available (Kaplan et al., 2021). This flexibility also affords the ability to manipulate aspects of the training content and environment so that trainees can better prepare for a wider range of conditions and scenarios they may experience outside of training. The amount of extraneous information within a simulated training environment can also be manipulated in order to expose trainees to fewer or greater sources of distractions (Thompson et al., 2009). Additionally, XR training solutions can be used for remote collaboration, accommodating training in situations where the trainees and instructors are not in the same physical space (Webel et al., 2013).

**Opportunities for Feedback, Practice, and Immersion.** Immediate feedback that can be offered by XR training promotes learning efficiency and accuracy (Kaplan et al., 2021). Training in XR environments can typically be paused to allow instructors to interject with additional guidance,
feedback, and demonstrations (Thompson et al., 2009). Furthermore, XR training solutions are likely repeatable, providing trainees with more opportunities to practice and reinforce their KSAs (Martin et al., 2014; Thompson et al., 2009). This repeatable trait also provides a consistent and standardized experience for all trainees (Martin et al., 2014). XR technologies can deliver a more immersive training experience. Higher immersion has been linked to more effective learning of procedural tasks (Morélot et al., 2021).

Specific to advantages of AR/MR training technologies, AR/MR users can access supplementary virtual information (e.g., instructions, images, videos, interface for authoring notes) while interacting with real-world objects (Neumann & Majoros, 1998; Sautter & Daling, 2021) that provide trainees with useful tactile feedback to be incorporated into their sensorimotor memory storage (Webel et al., 2013). Additionally, AR/MR technologies can annotate aspects of the physical world. Directing trainees’ attention in this way can reduce time spent searching for necessary information and minimize head movements (Marner et al., 2013; Polvi et al., 2018; Tang et al., 2003; Volmer et al., 2018), making training more efficient and decreasing expended physical effort. AR/MR annotations can also reduce the amount of attentional switching between the instructions and the equipment, resulting in decreased mental workload (Neumann & Majoros, 1998; Tang et al., 2003). Furthermore, because transfer of training depends on the physical similarity between the simulator and the real world (Hochmitz & Yuviler-Gavish, 2011) and AR/MR is closer to reality on the virtuality-reality spectrum, transfer of training may be better for AR/MR experiences compared to VR. AR/MR learning experiences can also facilitate trainees’ sense of presence, motivation to learn, engagement with the training content (Wu et al., 2013).

**Learning Theories Supporting Extended Reality Training**

One of the most compelling motivators for utilizing XR technologies in training is its ability to support experiential learning (Asad et al., 2021; Pomerantz, 2019). Experiential learning refers to “the
process whereby knowledge is created through the transformation of experience,” (Kolb, 1984, p. 41). Essentially, experiential learning is “learning by doing” (Gentry, 1990). Training and education courses can incorporate experiential learning through the implementation of active, participatory learning opportunities (Hawtrey, 2007).

Shown in Figure 6, the Experiential Learning Cycle is a widely-accepted framework for instructional design and curriculum development. This framework draws on the work of notable psychologists, philosophers, and educators, including John Dewey, Jean Piaget, and Kurt Lewin. The cycle assumes that knowledge results from grasping (i.e., taking in information) and transforming (i.e., interpreting and acting upon the grasped information) experience. Two modes of grasping experience are Concrete Experience and Abstract Conceptualization. Two modes of transforming experience are Reflective Observation and Active Experimentation (Kolb, 2014). As a whole, the Experiential Learning Cycle describes the process of learning as initiated by a Concrete Experience, followed by Reflective Observation about that experience, then Abstract Conceptualization to form conclusions about the meaning of the experience, and then engaging in Active Experimentation to apply what was learned from the experience (Institute of Experiential Learning, 2021).
XR training experiences afford trainees the opportunity to learn by doing. XR technologies enable experiential learning by providing concrete, hands-on experiences with concepts and activities that cannot be performed in the real-world due to lack of accessibility or safety (Asad et al., 2021; Pomerantz, 2019). XR can facilitate the ability to conceptualize abstract ideas by providing actual interactions with notions that were previously only theoretical. XR trainees are no longer limited to observational learning, as they can partake in realistic interactions that mimic real-world tasks (McGowin et al., 2021a).

Another theory that supports XR training is the situated learning theory, which emphasizes that what is learned is specific to the situation in which it is learned (Anderson et al., 1996). This notion implies that learning is enhanced when it takes place in the same context in which it is applied. Situated learning relates to transfer (i.e., the application of knowledge learned in one situation to another) and near-transfer (i.e., applying knowledge learned in one context to a similar context with only slight
An advantage of XR is that it can simulate real-world problems and contexts to provide near-transfer opportunities that enhance learning (Dunleavy & Dede, 2014).

A third idea that can be implemented within XR training experiences to promote learning is mnemonic devices. Mnemonic devices are memory techniques that can facilitate the transfer of information from short-term or working memory, to long-term memory (Sternberg & Sternberg, 2012). The Method of Loci is a mnemonic device where individuals walk around an area and link information to be remembered to specific landmarks within the area (Maguire et al., 2003; Sternberg & Sternberg, 2012). This mnemonic is based on the idea that people tend to memorize information more effectively when information can be tethered to a frame of reference in the real world, suggesting a strong relationship between spatial location, working memory, and long-term memory. AR/MR technologies can leverage this phenomenon by presenting virtual information that is linked to areas within the physical world (Tang et al., 2003). AR/MR training systems can reinforce this connection through repetition to increase the likelihood that trainees recall information linked to the real-world even without the presence of the virtual cue.

In summary, XR experiences have the potential to support learning by offering hands-on experiential learning opportunities that promote near-transfer and memory recall.

**Extended Reality Training Challenges**

While several advantages and supporting theories for XR training have been identified, there are key drawbacks, including performance and individual differences, physical discomfort, as well as cognitive, perceptual, and technical limitations.

**Performance and Individual Differences.** A trainee’s performance during simulation-based training may not reflect their operational performance. Trainees may experience different levels of stress during training and operational situations. Fatigue, complacency, and boredom tend to not be as prevalent during training, which may enhance performance during training. Trainees may also have
different expectations during training that may result in behavioral differences between training and operational environments. For example, unanticipated events may be expected during training, priming trainees to prepare for quick and accurate reactions to such events. Additionally, trainees would likely have a more recent review of the intended KSAs that they could more readily apply to their training performance. However, such a recent review is less likely to occur in operational settings, where potential memory degradation of KSAs may impact performance (Thompson et al., 2009). As for individual differences, XR may not be a suitable training delivery method for all trainees. Trainee learning styles, acceptance of XR technologies, motivation to learn, self-efficacy, and spatial abilities can impact their ability to learn in XR training environments (Ling et al., 2021; Peracchio, 2020; Sytwu & Wang, 2015; Taylor et al., 2022).

Focusing on spatial abilities (i.e., the ability to understand, generate, and transform visual images; Lohman, 1996), prior research has found significant positive correlations between spatial ability and VR training performance, such that higher spatial ability improves performance (Hamblin, 2005; Peracchio, 2020). Additionally, spatial ability may be effective in predicting training performance measures, such as learning efficiency and transfer of training (Hamblin, 2005), as well as in predicting trainee affective (i.e., satisfaction with training) and utility (i.e., perceived practical value of the training) reactions (Peracchio, 2020). Common measures for spatial ability include assessments that require participants to mentally rotate or transform objects (Lohman, 1996), such as the Ekstrom et al. (1976) Paper Folding Test (PFT). The PFT measures spatial visualization (i.e., an individual’s ability to mentally manipulate two- and three-dimensional figures), a factor of spatial ability (McGee, 1979). Spatial visualization is an individual difference that has been shown to vary by age and gender, with males typically scoring higher (Goldstein et al., 1990) and older adults typically scoring lower (Salthouse et al., 1990).
Physical Discomfort. Use of XR systems may cause discomfort. Cybersickness (i.e., adverse side effects, such as dizziness, nausea, drowsiness, and visual stress, following exposure to XR systems) is a widely observed condition in XR users. Cybersickness symptoms tend to be less severe in AR/MR users compared to VR users, likely due to the fact that AR/MR users can still see their physical surroundings and therefore experience less of the visual-vestibular mismatch that often causes motion-sickness (Stanney et al., 2021). A recent medical case study attributed a patient’s cervical spine injury to prolonged use of a VR headset. The authors concluded that rapid, repetitive movements within the VR gaming environment coupled with the weight of the VR headset contributed to a stress fracture (Baur et al., 2021). As the adoption of XR devices increases, it is important to determine and implement approaches for minimizing physical discomfort and injury.

Cognitive Limitations. Trainees may become reliant on virtual elements (e.g., instructions, annotations) presented by AR/MR systems (Tang et al., 2003; Webel et al., 2013), negatively impacting learning transfer if trainees are required to complete the task without the assistance of virtual content. This may lead to trainees simply following the virtual instructions instead of taking the time to understand the procedure. Over-reliance can also lead to too much trust in the virtual content, preventing users from switching their attention to important real-world cues when needed (Gabbard et al., 2014; Tang et al., 2003). The amount of virtual elements matters, as well. Displaying too much virtual content can result in a cluttered view of the real-world and cause feelings of informational overload in users (Gabbard et al., 2014).

Perceptual Limitations. AR/MR systems have potential problems with perception and legibility of virtual elements, as AR/MR devices must display virtual content on top of a vast range of objects and backgrounds (Hillmann, 2021; Kruijff et al., 2010). Currently, known capabilities of AR/MR technologies do not include real-time modifications that adapt the appearance of virtual content to ensure legibility against different elements in the physical world. Other perceptual issues of AR/MR technologies include...
latency (i.e., delays in the display of virtual content), vergence-accommodation conflict (i.e., when the user’s eyes converge on the virtual content that appears to be farther away in the physical environment, but the eye’s lenses accommodate to the closer display screen), field of view (FOV) limitations, and distortion of colors presented in the real world (Azuma, 1997; Kruijff et al., 2010; Livingston et al., 2012). Furthermore, virtual content display by AR/MR devices can occlude (i.e., block from view) important information in the real world to which the user needs to attend (Gabbard et al., 2014; Kruijff et al., 2010).

Technical Limitations. The effectiveness of training with XR technologies may be restricted by the devices’ technical capabilities. Registration errors (i.e., when virtual content does not properly align with the physical world) due to limitations in AR/MR device tracking capabilities can result in virtual elements appearing in improper locations and can compromise the illusion that the virtual and real worlds seamlessly coexist (Azuma, 1997; Gabbard et al., 2014; Kruijff et al., 2010; Livingston et al., 2012). Other technical limitations include the rendering quality of virtual content, as well as device-related issues such as weight and battery capacity. Problems with battery capacity can be a compounding issue that places constraints on the use of wireless devices, causing users to either be tethered to a system or to carry bulky equipment (Goh et al., 2021; Minna et al., 2021). Both of these workarounds can pose safety risks to users, as tethered systems present tripping hazards and bulky equipment can cause physical discomfort. Many issues related to technical limitations are anticipated to resolve with future advances in XR hardware and software capabilities (Goh et al., 2021).

Extended Reality Training Effectiveness

Several review papers published within the past few years have examined the use of XR technologies in training and education, drawing conclusions as to whether these technologies are effective in promoting learning and performance outcomes. Table 1 summarizes the findings of nineteen recent review papers, sorted first by AR, VR, and XR, then alphabetically within each of the three
categories. Of these review papers, there are three meta-analyses. Eight review papers were broader in scope and included training and education methods for a variety of disciplines. Six papers focused on the medical discipline, each of which further concentrated on a medical subspecialty.
Table 1

Summary of Recent XR Training and Education Review Papers

<table>
<thead>
<tr>
<th>Study</th>
<th>Paper Title</th>
<th>Review Type</th>
<th>XR/VR/AR/MR</th>
<th>Discipline</th>
<th># of Included Papers</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzahrani (2020)</td>
<td>Augmented Reality: A Systematic Review of Its Benefits and Challenges in E-learning Contexts</td>
<td>Systematic Review</td>
<td>AR</td>
<td>Multiple; Education</td>
<td>28</td>
<td>Benefits of AR in e-learning include enhanced student engagement, motivation, attention/focus, and knowledge retention. Challenges include information and cognitive overload, lack of experience in using the technology, and technical issues.</td>
</tr>
<tr>
<td>Han et al. (2022)</td>
<td>Augmented Reality in Professional Training: A Review of the Literature from 2001 to 2020</td>
<td>Systematic Review and Meta-Analysis</td>
<td>AR</td>
<td>Multiple; Training</td>
<td>49</td>
<td>Of the fifteen included AR effectiveness experimental studies, nine showed positive effects, five found insignificantly negative effects, and one concluded no effect. There was an overall small positive effect size of AR-supported instruction on learning outcomes.</td>
</tr>
<tr>
<td>Checa &amp; Bustillo (2020)</td>
<td>A Review of Immersive Virtual Reality Serious Games to Enhance Learning and Training</td>
<td>Systematic Review</td>
<td>VR</td>
<td>Multiple; Training &amp; Education</td>
<td>68 Education; 67 Training</td>
<td>VR demonstrably enhanced learning in 30% of the included education papers, and 29% of the included training papers.</td>
</tr>
<tr>
<td>Hamilton et al. (2021)</td>
<td>Immersive Virtual Reality as a Pedagogical Tool in Education: A Systematic Literature Review of Quantitative Learning Outcomes and Experimental Design</td>
<td>Systematic Review</td>
<td>VR</td>
<td>Multiple; Education</td>
<td>29</td>
<td>About half of the included studies found a significant advantage of utilizing VR in education over less immersive learning methods, particularly for abstract subjects and procedural tasks. Many studies found no significant benefit of using VR over less immersive technology. Two studies found detrimental effects of VR.</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Title</td>
<td>Type</td>
<td>Technology</td>
<td>Sample Size</td>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Portelli et al. (2020)</td>
<td>Virtual Reality Training Compared with Apprenticeship Training in Laparoscopic Surgery: A Meta-Analysis</td>
<td>Meta-Analysis</td>
<td>VR</td>
<td>24</td>
<td>VR improves efficiency and quality (e.g., reduced error rate, improved tissue handling) of trainees’ surgical practice. However, current VR capabilities fail to produce valuable haptic feedback to trainees.</td>
<td></td>
</tr>
<tr>
<td>Arjomandi Rad et al. (2021)</td>
<td>Extended, Virtual and Augmented Reality in Thoracic Surgery: A Systematic Review</td>
<td>Systematic Review</td>
<td>XR</td>
<td>21</td>
<td>Of the seven included training effectiveness studies, junior trainees benefit most from simulation-based training.</td>
<td></td>
</tr>
<tr>
<td>Cross et al. (2022)</td>
<td>Using Extended Reality in Flight Simulators: A Literature Review</td>
<td>Systematic Review</td>
<td>XR</td>
<td>39</td>
<td>Focused on VR systems. More research is required to determine if VR can successfully replace traditional flight simulators.</td>
<td></td>
</tr>
<tr>
<td>Dadario et al. (2021)</td>
<td>Examining the Benefits of Extended Reality in Neurosurgery: A Systematic Review</td>
<td>Systematic Review</td>
<td>XR</td>
<td>116</td>
<td>AR technology is considered more developed for surgical training compared to VR. Improvements were demonstrated in resident knowledge and performance in 64% of the 33 resident studies. However, expense of XR technologies prevents adoption.</td>
<td></td>
</tr>
<tr>
<td>Doolani et al. (2020)</td>
<td>A Review of Extended Reality (XR) Technologies for Manufacturing Training</td>
<td>Systematic Review</td>
<td>XR</td>
<td>52</td>
<td>XR training can have a positive impact on manufacturing training. Immersive XR training promotes performance and increases engagement.</td>
<td></td>
</tr>
<tr>
<td><strong>Le Noury et al. (2022)</strong></td>
<td>A Narrative Review of the Current State of Extended Reality Technology and How it can be Utilised in Sport</td>
<td>Narrative Review</td>
<td>XR</td>
<td>Sports Training</td>
<td>N/A</td>
<td>Despite a lack of research on XR usage in sports, XR may be a promising tool for sports training, particularly for improving perceptual-cognitive skills.</td>
</tr>
<tr>
<td><strong>Longo et al. (2021)</strong></td>
<td>Augmented Reality, Virtual Reality and Artificial Intelligence in Orthopedic Surgery: A Systematic Review</td>
<td>Systematic Review</td>
<td>XR</td>
<td>Medical; Orthopedic Surgery</td>
<td>21</td>
<td>Of the four included training effectiveness studies, all four utilized VR technologies. VR training was concluded to have significant benefits (e.g., improved skills, fewer errors during surgery, faster surgery performance) compared to traditional training methods.</td>
</tr>
<tr>
<td><strong>Maas &amp; Hughes (2020)</strong></td>
<td>Virtual, Augmented, and Mixed Reality in K–12 Education: A Review of the Literature</td>
<td>Narrative Review</td>
<td>XR</td>
<td>Multiple; K-12 Education</td>
<td>29</td>
<td>Of the 11 included studies that measured performance/learning outcomes, use of XR produced better outcomes in 6 studies. No difference was found in 3 studies, and 2 studies found that XR produced worse outcomes due to issues with usability and distractibility.</td>
</tr>
<tr>
<td><strong>Minna et al. (2021)</strong></td>
<td>A Systematic Literature Review on Extended Reality: Virtual, Augmented and Mixed Reality in Collaborative Working Life Setting</td>
<td>Systematic Review</td>
<td>XR</td>
<td>Multiple; Training</td>
<td>26</td>
<td>Inconsistent findings across the five included XR training effectiveness papers.</td>
</tr>
<tr>
<td><strong>Ong et al. (2021)</strong></td>
<td>Applications of Extended Reality in Ophthalmology: Systematic Review</td>
<td>Systematic Review</td>
<td>XR</td>
<td>Medical; Ophthalmology</td>
<td>87</td>
<td>Of the 4 included randomized trial studies that compared VR with conventional training methods, three showed significant performance improvements (e.g., quality, efficiency, efficacy, completion rates) in favor of VR training.</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Title</td>
<td>Method</td>
<td>XR Application</td>
<td>Literature Count</td>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>--------</td>
<td>----------------</td>
<td>------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Taylor et al. (2022)</td>
<td>Extended Reality Anatomy Undergraduate Teaching: A Literature Review on an Alternative Method of Learning</td>
<td>Systematic Review</td>
<td>XR</td>
<td>Medical; Undergraduate Anatomy Education</td>
<td>45</td>
<td>Compared to traditional teaching methods, eight studies showed improved effectiveness of AR or VR. Both AR and VR had high rates of satisfaction and acceptability as a teaching aid.</td>
</tr>
</tbody>
</table>
Across all reviews, the majority of included studies showed positive impacts of XR technologies on learning and performance outcomes. A number of included papers found no significant difference between XR and traditional training methods, suggesting equivalent effectiveness. Only a small number of included papers found negative impacts on learning and performance. Most of the included papers investigated VR devices, indicating a need for further insight into AR/MR training effectiveness. One limitation of these reviews is that they only include peer-reviewed publications. However, there are likely many other reports that investigate AR/MR training effectiveness that may not be published due to use of proprietary technology, inclusion of sensitive information associated with the tasks being trained, or lack of motivation in publishing insignificant or negative findings.

Multiple review articles itemized the metrics used by the included papers to assess training effectiveness. In general, pre- and post-tests, interviews, and questionnaires were commonly used to measure changes in learner outcomes (Maas & Hughes, 2020). Commonly used objective metrics consisted of precision, accuracy, awareness, reaction time, error rate, and time to completion. Frequently used subjective metrics included engagement, enjoyability, and user feedback (Doolani et al., 2020; Han et al., 2022).

**Current Literature Gaps**

Kaplan et al. (2021) noted in their meta-analysis of XR training effectiveness that current literature was sparse. In particular, the number of studies investigating AR/MR technologies was insufficient for some analyses. The authors also found that while XR may be more suitable for physical tasks (e.g., learn a spatial, procedural task) compared to cognitive tasks (e.g., memorizing facts about plants), inconsistent findings among the included studies make it difficult to conclude which types of tasks are more amenable for XR training (Kaplan et al., 2021).

Other reviews also indicate a higher prevalence of VR training effectiveness investigations (Minna et al., 2021) and a lack of AR/MR inquiries (Han et al., 2022). While this may be an artifact of the
greater ubiquity of VR technologies (Hillmann, 2021), the anticipated growth of the AR/MR market (Fortune Business Insights, 2022) and a lack of understanding regarding limitations unique to AR/MR warrant the need for additional investigations of AR/MR training effectiveness. Additionally, AR/MR training effectiveness review paper authors have found that handheld AR/MR devices (e.g., tablets, smartphone) and desktop/laptop computers have been investigated more frequently (Han et al., 2022; Werrlich et al., 2017). Therefore, further research is needed to understand how AR/MR training experiences delivered through HMDs impact training effectiveness. Additionally, Werrlich et al. (2017) calls upon future studies to investigate the impact of AR/MR HMD training on short- and long-term knowledge retention and performance.

Specific to the type of task being trained, there are a number of studies that investigate AR/MR training effectiveness of procedural tasks. However, the majority of this existing research focuses on very specific procedural tasks that require domain-specific expertise (e.g., Bifulco et al., 2014; Chen & Liao, Jul 2015; Gonzalez-Franco et al., 2017) and may not generalize to other procedural tasks or populations. Additionally, some of these studies utilized outdated devices (e.g., Kolla et al., 2021; Werrlich et al., 2018) or took place several years ago (e.g., Henderson & Feiner, Oct 2009; Tang et al., 2003; Valimont et al., 2007). As a result, their findings may not accurately reflect technological advances seen in today’s AR/MR devices.

**Purpose**

Significant investments are being poured into XR and XR training solutions by several organizations around the world, and future forecasts show no sign of these investments slowing down. However, prior research has a limited understanding as to whether AR/MR delivered training experiences are effective. It is especially unclear as to whether modern AR/MR HMDs are suitable for training procedural tasks and whether they can promote long-term knowledge retention. This study aims to investigate short- and long-term training effectiveness of AR/MR training delivered through a
HMD for a procedural task. Results from this research intend to provide insight into the implications of adopting AR/MR HMDs for procedural task training.

**Research Design**

The purpose of this study is to investigate training effectiveness of AR/MR instructions delivered through a HMD for a procedural task. This study utilized a within-subjects 3x3 experimental design. The independent variables were time of procedure recall (immediate vs. post-test vs. retention-test) and instruction method (paper vs. AR (images only) vs. MR (images and virtual cues)). Table 2 portrays the nine study conditions: 1) paper / immediate recall, 2) AR / immediate recall, 3) MR / immediate recall, 4) paper / post-test, 5) AR / post-test, 6) MR / post-test, 7) paper / retention-test, 8) AR / retention-test, 9) MR / retention-test. Within groups comparisons were conducted across instruction methods to compare measures between training delivered through paper, AR, and MR instructions. Within groups comparisons across time of recall were also performed. Additionally, interaction effects were analyzed to determine if instruction method impacts measures over time.

**Table 2**

*Study Conditions*

<table>
<thead>
<tr>
<th>Instruction Method</th>
<th>Time of Procedure Recall</th>
<th></th>
<th>Retention-Test (1 week later)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Immediate</td>
<td>Post-Test</td>
<td></td>
</tr>
<tr>
<td>Paper</td>
<td>Condition 1</td>
<td>Condition 4</td>
<td>Condition 7</td>
</tr>
<tr>
<td>AR (images only)</td>
<td>Condition 2</td>
<td>Condition 5</td>
<td>Condition 8</td>
</tr>
<tr>
<td>MR (images and virtual cues)</td>
<td>Condition 3</td>
<td>Condition 6</td>
<td>Condition 9</td>
</tr>
</tbody>
</table>
**Hypotheses**

**Performance**

The Ganier et al. (2000) Model of Processing Procedural Instructions provides insight into how trainees understand instructions. Because the MR training will present virtual cues that display the procedure instructions onto the participants’ workspace, it may be easier for participants to match their actions with those outlined in the instructions, leading to a more accurate understanding of the procedure. Participants in this study will likely be less familiar with AR-/MR-based instructions (Albaladejo et al., 2020; Hillmann, 2021; Werrich et al., 2018). Better familiarity with the paper instructions may positively impact performance because participants may be able to focus more on learning the procedure instead of learning how to use the instruction method. Therefore, it is hypothesized that performance will be higher for paper and MR compared to AR. Additionally, loss of trained knowledge and skills can occur over time (Arthur et al., 1998). As a result, it is also hypothesized that performance will be highest for immediate recall, less for post-test recall, and lowest for retention recall. Finally, because AR lacks virtual cues and is expected to be unfamiliar to participants, it is hypothesized that AR will have the biggest decrease in performance over time.

**H1.** Performance of the paper folding procedural task will be influenced by instruction method in that paper and MR will result in higher accuracy than AR.

Performance between instruction methods: Paper, MR > AR

**H2.** Performance of the paper folding procedural task will be influenced by time of recall in that immediate recall will result in higher accuracy than post-test, followed by retention-test. The difference in performance between immediate recall and retention-test will be greater for the AR condition.

Performance over time: Immediate > Post > Retention

Interaction: AR will have biggest decrease in performance over time
**Time to Completion**

Similar to the performance hypotheses, paper and MR instructions may result in a better understanding of the procedure, resulting in faster completion times during procedure recall. Thus, it is hypothesized that time to completion will be faster for paper and MR, and slower for AR. Additionally, because knowledge and skill decay is expected to occur over time, participants may utilize additional time to recall the procedure as time since training increases. Therefore, it is also hypothesized that time to completion will be fastest for immediate recall, slower for post-test recall, and slowest for retention recall.

**H3.** Time to complete the paper folding procedural task will be influenced by instruction method in that paper and MR will result in faster completion times than AR.

- Time to completion between instruction methods: Paper, MR < AR

**H4.** Time to complete the paper folding procedural task will be influenced by time of recall in that retention-test will take more time to complete than post-test, followed by immediate recall.

- Time to completion over time: Retention > Post > Immediate

**Perceived Difficulty**

As participants are less likely to be familiar with AR/MR training, it is hypothesized that they will perceive the AR and MR training experiences as more difficult than the paper condition. Additionally, anticipated knowledge and skill decay may result in participants finding recall more difficult over time. So, it is also hypothesized that retention recall will be perceived as more difficult, post recall as less difficult, and immediate recall as least difficult.

**H5.** Perceived difficulty of the training experience will be influenced by instruction method in that AR and MR will result in higher perceived difficulty than paper.

- Perceived difficulty between instruction methods: AR, MR > Paper
**H6.** Perceived difficulty of the paper folding procedural task will be influenced by time of recall in that retention-test will result in higher difficulty than post-test, followed by immediate recall.

Perceived task difficulty over time: Retention > Post > Immediate

**Perceived Confidence**

Higher self-efficacy, or trainees’ confidence in their ability to successfully complete a task, can promote learning (Bisbey et al., 2021; Salas et al., 2012). Confidence in one’s ability to successfully complete a task can be influenced by prior experience (Davis et al., 2000), such that more experience or familiarity with the task and equipment used to complete the task can increase self-efficacy (Margolis & McCabe, 2006; Prieto & Altmair, 1994). Since participants in this study will likely be less familiar with AR-/MR-based instructions, it is hypothesized that they will have less perceived confidence in accurately performing the procedural task using AR/MR instructions compared to paper instructions. Additionally, based on prior work that indicates confidence in completing procedural tasks declines as time since training increases (Buttussi & Chittaro, 2021; Schumann et al., 2012), it is also hypothesized that confidence will be highest for immediate recall, less for post-test recall, and lowest for retention recall.

**H7.** Confidence associated with the paper folding procedural task will be influenced by instruction method in that AR and MR will result in lower confidence than paper.

Perceived confidence between instruction methods: AR, MR < Paper

**H8.** Confidence associated with the paper folding procedural task will be influenced by time of recall in that retention-test will result in lower confidence than post-test, followed by immediate recall.

Perceived confidence between over time: Immediate > Post > Retention

**Workload**

Applying the Ganier et al. (2000) Model of Processing Procedural Instructions to AR/MR training suggests that workload will increase during the AR condition because of an increased amount of
attentional switching between the instructions and the participants’ workspace. The model also indicates that workload during the MR condition will decrease because the MR virtual cues serve to integrate the instructions within the participants’ workspace, reducing the amount of attentional switching required to complete the task. However, because participants are expected to be less familiar with AR-/MR-based instructions, they will likely have to expend additional resources in these conditions as they juggle learning how to use the AR/MR device while learning the procedure. As a result, it is hypothesized that participants will report higher workload for AR, less workload for MR, and least workload for paper.

**H9.** Workload associated with the paper folding procedural task will be influenced by instruction method in that AR will result in higher workload MR, followed by paper.

  Workload: AR > MR > Paper

*User Experience*

Technical limitations of modern AR/MR experiences (e.g., registration errors, occlusion, latency) are expected to have a negative impact on participants’ perception of AR/MR utility. Between AR and MR, the presence of the MR virtual cues may be perceived as more useful. Therefore, it is hypothesized that paper will be perceived as more straightforward and practical compared to AR and MR, resulting in highest pragmatic user experience scores for paper, lower for MR, then lowest for AR. Similar to findings presented by Werrich et al. (2018), it is also hypothesized that MR and AR will be perceived as more novel and innovative compared to paper, resulting in highest hedonic user experience scores for MR, lower for AR, then lowest for paper.

**H10.** User experience associated with the paper folding procedural task will be influenced by instruction method in that paper will result in a higher pragmatic score than MR, followed by AR.

  Pragmatic: Paper > MR > AR
**H11.** User experience associated with the paper folding procedural task will be influenced by instruction method in that MR will result in a higher hedonic score than AR, followed by paper.

Hedonic: MR > AR > Paper

*Trainee Reactions*

Similar to the user experience hypotheses, AR/MR technical limitations and expected lack of familiarity with AR/MR devices may result in more positive trainee reactions for paper-based instructions. MR may be viewed more favorably than AR due to the possibility of the virtual cues being perceived as useful. As a result, it is hypothesized that trainee reactions will be most positive for paper, less positive for MR, then least positive for AR.

**H12.** Trainee reactions associated with the paper folding procedural task will be influenced by instruction method in that paper will result in higher trainee reactions than MR, followed by AR.

Trainee Reactions: Paper > MR > AR

*Exploratory Hypotheses*

Additional analyses that will be investigated include:

**H13.** Relationships between select demographic variables (e.g., spatial visualization abilities; prior XR, video game, and origami experience) and performance will be examined.

**H14.** Differences between the dependent variables (e.g., time to completion, perceived task difficulty, perceived confidence) for paper, AR, and MR first and second training sessions will be examined.

**H15.** Differences in strategies for completing the paper folding procedural task between instruction methods will be examined.

**H16.** Differences in likes, dislikes, and recommendations for improvement between paper, AR, and MR will be examined.
Method

Participants

Participants were recruited from Embry-Riddle Aeronautical University (ERAU; Daytona Beach), a private university located in the southeastern United States, as well as from the surrounding community. In order to be eligible to participate in the study, participants must have been eighteen years old or older, have had normal or corrected-to-normal vision, and have had full use of both hands and arms. Participants were recruited through word-of-mouth, email, and social media.

Prior to recruiting participants for data collection, an application was submitted to ERAU’s Institutional Review Board (IRB). The IRB reviewed the study to ensure the rights and welfare of participants were protected before, during, and after data collection. The researchers involved in the study also completed training related to background, principles, and regulations associated with the protection of human research participants. Participation in the study was voluntary and could be discontinued at any time during the study without negative consequences to the participant.

Participants were required to review and sign an informed consent document prior to their study participation. The informed consent document detailed the study’s purpose, reasonably foreseeable risks or discomforts, potential benefits, procedure for maintaining confidential records of personally identifiable information, and researcher contact information.

Materials

Task

Participants completed a series of origami models, following instructions delivered via paper or through the Microsoft HoloLens 2. Six different models were selected to be included in this study to prevent learning effects that may impact the study’s results if the same model was used for all instruction methods. Three models were used as practice models to familiarize participants with the diagram notations (e.g., solid lines, dashed lines, arrows, grid) and instruction delivery method (i.e.,
paper, AR, MR): Cup, Dog, and Sailboat. The three training and experimental models were the Cicada, Samurai Helmet, and Necktie. These models were chosen because they have a similar number of steps and the same types of folds. They were also perceived to be of comparable difficulty by pilot participants, minimizing variation in the dependent variables that may result in using different models for each instruction method. Diagrams for all six models were adapted from origami instructions available online (see Appendix A). Participants completed each model using a square piece of paper measuring 8.5 by 8.5 inches.

Origami model instructions were presented to participants one step at a time, regardless of instruction method. Paper instructions were presented by printing each step diagram onto separate cards measuring 8.5 by 5.5 inches. These cards were placed in sheet protectors within a binder. AR instructions were presented through the HoloLens 2. MR instructions were also presented through the HoloLens 2, but were supplemented by virtual cues anchored to the participant’s workspace. The virtual content included arrows, dashed lines, and text in specific locations to mimic the notations presented on the paper and digital diagrams. Virtual content was designed for maximum visibility by using sizes and colors that would stand out against the participants’ physical workspace. See Figures 7 and 8 for images of how the paper, AR, and MR instructions were displayed to participants.
Figure 7

Appearance of Paper Instructions

Paper - Step 1 of Cup

Paper - Step 2 of Cup
Figure 8

Appearance of AR and MR Instructions

AR - Step 1 of Cup

AR - Step 2 of Cup

MR - Step 1 of Cup

MR - Step 1 of Cup
**Device**

Shown in Figure 9, the Microsoft HoloLens 2 is a wireless HMD first released in 2019 that has since become the world’s leading AR/MR HMD (Pu et al., 2022). The device uses spatial mapping technology to construct three-dimensional (3D) models of a user’s physical surroundings. Virtual elements can be displayed on top of the user’s real-world environment, and even anchored to physical objects or surfaces. Users can interact with digital content through hand tracking, eye tracking, and voice commands (Microsoft, 2022a). The Microsoft HoloLens 2 was used in the AR and MR training conditions.

**Figure 9**

*Microsoft HoloLens 2*
**Application**

During the AR and MR training conditions, the Microsoft HoloLens 2 was used to launch an application called Microsoft Dynamics 365 Guides (see Figure 10). Dynamics 365 Guides supports two user types: Authors and Operators. Authors can use the HoloLens 2 application and its affiliated personal computer (PC) application to create AR/MR instructional guides composed of text, images, videos, and 3D virtual objects. These 3D virtual objects can be anchored to specific locations within an Operator’s physical surroundings. Because Dynamics 365 Guides is a marker-based application, 3D virtual objects are anchored in relation to a quick response (QR) code that is displayed in a location central to the Operator’s work area.

The researcher used the Author mode to create the AR and MR environments. Twelve instructional guides were created in the PC application to produce an AR and MR variant for each practice and experimental model (e.g., Cup AR, Cup MR; Necktie AR, Necktie MR). Diagrams for each step of a model were transformed into .JPG images and uploaded to their respective guide. The MR variants for each model contained 3D virtual cues that mirrored notations presented in the diagrams (e.g., dashed lines, arrows, step number, text instructions). 3D arrows and step numbers were taken from the Dynamics 365 Guides asset library. 3D models for the dashed lines and text instructions were created using Tinkercad, a free online 3D modeling software program (Autodesk, 2023), and saved as .GLB files. These files were uploaded to their respective guides using the Dynamics 365 Guides PC application. Once all of the files were uploaded to each guide, the HoloLens 2 application was used to resize, place, and anchor MR virtual cues to the participants’ workspace.

Once an Author completes an instructional guide, Operators can use the guide to finish a task or set of tasks by paginating through a series of step cards. Operators will see only one step card at a time. Operators can choose to advance to the next step, or return to a previous step, by using their hand to point to the next/back arrows. If an Operator prefers to have hands-free control of the interface, they
can opt to use voice commands or the eye tracking input method by holding their gaze within the confines of the arrow buttons until the interface advances to the next step. Operators can also choose to have the step cards follow them as they move about their physical space, or to lock the step cards in a particular location of their choice (Microsoft, 2022b).

**Figure 10**

*Microsoft Dynamics 365 Guides*


**Study Station**

Participants completed origami models in all three instruction method conditions while seated in front of a table-top study station. As shown in Figure 11, the study station consisted of a grid measuring 12 by 12 inches (individual grid squares measured 2 by 2 inches) and a QR code. A piece of
plexiglass measuring 16 inches by 20 inches was placed over the grid to promote durability of the study station and to affix the QR code to ensure it stayed the same distance away from the grid for all participants. For all instruction methods, participants were prompted to fold the origami model on the grid surface and to align their paper to the grid as they complete the model. The QR code was only used in the AR and MR conditions. The QR code was scanned by participants after they put on the HoloLens 2 so that the virtual content appeared in relation to the QR code and study station.

Prior to data collection, the procedure and all study materials were piloted by three participants, whose data is not included in this study. These pilot participants were helpful in confirming the accuracy of the diagrams, the visibility of the virtual content against the physical workspace, and the flow of the study’s procedure. Pilot participants also assisted with the selection of the origami models used in this study by completing origami models with a comparable number of steps and rating how difficult or easy it was to complete each model. Models rated similarly in terms of difficulty were chosen for this study.
Figure 11

Study Station
Measures

Demographics

The demographic questionnaire (see Appendix B) was distributed at the beginning of the study using Qualtrics, an online survey platform. The questionnaire collected basic information about the participants, including their age, gender, highest level of completed education, and occupation. Questions related to experience with and ownership of XR devices were also included. Additionally, the demographics questionnaire collected participants’ video game experience, as well as familiarity and experience with origami.

Paper Folding Test

Following the demographics questionnaire, participants completed the Paper Folding Test (PFT; see Appendix C). The PFT is composed of two parts, each with ten questions that participants must complete within three minutes. Each question consists of presenting participants with a series of diagrams that represent a square piece of paper being folded one to three times before a hole is punched through the folded paper. Participants must choose between five diagrams what the paper and hole pattern would look like after the paper is unfolded (Ekstrom et al., 1976). The PFT was scored by counting the number of correct answers out of all twenty questions, as well as by calculating the proportion of correct to attempted answers (Ekstrom et al., 1976; Jaeger, 2015).

Performance

Each participant produced 18 origami models, all of which were organized and kept in order to measure performance after data collection was completed. To assess performance, researchers created scoring rubrics (see Appendix D) that evaluated the accuracy of participants’ completed Cicada, Samurai Helmet, and Necktie origami models. Each rubric was piloted multiple times with a variety of low- to high-quality models to ensure the rubric was comprehensive and accommodating to a range of model attempts. Three raters were trained on how to use the rubric. Rubric training involved teaching the
raters how to complete each origami model, presenting them with examples of pass/fail models for each rubric item, and rating at least five low- to high-quality models for each of the three origami patterns. Feedback was provided on these training ratings. After learning the rubrics, raters scored all of the participants’ Training 2, Immediate, Post-Test, and Retention attempts for the Cicada, Samurai Helmet, and Necktie.

**Time to Completion**

Participants trained on each origami model by completing it twice with the instructions accessible, before recalling the model three times at different time periods without instructions or assistance. The amount of time it took participants to complete each training (with instructions) and recall (without instructions) model was measured in seconds. The timer started when the participant indicated they understood what was expected of them and that they were ready to begin the task. The timer stopped when the participant indicated they completed the task.

**Difficulty and Confidence**

After participants completed each origami model, they were asked to fill out the Single Ease Question (see Appendix E) as a measure of task difficulty. The Single Ease Question is a one-item self-report as to how easy or difficult the participant found the task. Participants were also asked to fill out an additional item (see Appendix F) that aimed to capture participants’ confidence as to whether they successfully completed the origami model. Following their completion of each instruction method, participants were asked to complete a third item to indicate how easy or difficult it was for them to use that instruction method (see Appendix E). One-item Likert questions have been found to be easily administered with comparable sensitivity to more intricate measures (Sauro & Dumas, 2009).

**Workload**

Workload was measured three times, following exposure to each instruction method, using a modified version of the National Aeronautics and Space Administration Task Load Index (NASA-TLX; Hart
Staveland, 1988) called the Raw TLX (RTLX; Hart, 2006). The RTLX consists of six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration. Each subscale comprises one item rated on a line with 21 vertical tick marks (see Appendix G). Higher ratings indicate that participants perceived the task as more demanding or that they performed poorly. Both overall and subscale ratings were calculated by averaging the ratings to generate subscale and overall workload scores.

**User Experience**

User experience was measured three times, following exposure to each instruction method, using the short version of the User Experience Questionnaire (UEQ-S; Schrepp et al., 2017). The UEQ-S is composed of eight items (see Appendix H) providing insight into two dimensions: pragmatic quality and hedonic quality. Pragmatic quality (see first four items in Appendix H) refers to aspects related to the user’s task or goals, such as efficiency and clarity. Hedonic quality (see last four items in Appendix H) refers to aspects not related to the user’s task or goals, such as pleasure. The UEQ-S is recommended for use in experimental settings where participants are asked to assess the user experience of several products or variants of a product during one session, as was done in this study. Ratings for the first four items were averaged to produce a rating for the pragmatic quality scale, while the last four items were averaged to provide a score for the hedonic rating scale. All eight items were averaged to produce an overall user experience score (Schrepp et al., 2017).

**Trainee Reactions**

Trainee reactions were measured three times, following exposure to each instruction method, using six questions adapted from Long et al. (2008) and three open-ended response questions (see Appendix I). The six items from Long et al. (2008) cover three dimensions: technology satisfaction, enjoyment, and relevance of course content. Participants responded to each item using a five-point Likert scale (1 = Strongly Disagree; 5 = Strongly Agree). The three open-ended response questions aimed
to collect qualitative data regarding participant likes and dislikes about each training method, as well as recommendations they have for improving each training method.

**Cybersickness**

Cybersickness (i.e., adverse side effects, such as dizziness, nausea, drowsiness, and visual stress, following exposure to XR systems) was measured at the end of the first study session using the Simulator Sickness Questionnaire (SSQ; Kennedy et al., 1993). The SSQ consists of sixteen symptoms that are divided into three subscales: nausea, oculomotor, and disorientation. Participants rate the severity of their post-simulator experience with each symptom using a four-point scale (None, Slight, Moderate, Severe). Both overall and subscale ratings were computed using the formulas detailed in Kennedy et al. (1993).

**Post- and Retention-Test Open-Ended Questions**

At the end of the first (post-test) and second study session (retention-test), participants were asked a series of open-ended questions to collect their perceptions of the training experiences and the strategies they used to perform the procedure. The following questions were asked post-test as well as after the retention-test:

1. Please rank your preference for instructions provided using paper, AR, or MR, from most to least preferred. Why did you list the instruction methods in this order?
2. If you have to choose between instructions provided using AR or MR, which one would you choose? Why?

In addition, the following questions were asked after the retention-test:

1. What strategies did you use to recall the procedure for each origami model?
2. Do you think the method of instruction affected your ability to recall each origami model today?
Procedure

When participants arrived at the study location, they were given an overview of the study and prompted to review the informed consent document. After providing their consent, participants filled out the demographic questionnaire and completed the spatial ability test (PFT). The researcher then provided an overview of origami and explained how to interpret origami diagrams.

Participants then completed the first instruction method condition. The three instruction method conditions were counterbalanced across all participants. Paper instructions were delivered using a binder that held step cards measuring 8.5 by 5.5 inches. AR and MR instructions were delivered using the Microsoft Dynamic Guides 365 application on the Microsoft HoloLens 2. Before putting on the HoloLens 2, participants were given an overview of how to wear and adjust the headset, including how to tighten and loosen the headband and overhead strap, as well as how to adjust the brightness of virtual elements. Once the HoloLens 2 was fitted comfortably, participants completed the device’s eye calibration procedure and learned how to interact with virtual content using gestures and voice commands.

For each instruction method condition, participants completed a practice model (Cup, Dog, or Sailboat). The purpose of completing the practice model was to ensure participants were familiar with the diagram notations and the process of following the paper-, AR-, and MR-delivered instructions. Participants were prompted to fold using “hard” creases and to make sure their model aligned with the grid as shown in the diagram before moving onto the next step. During each instruction condition (Cicada, Samurai Helmet, and Necktie) participants practiced each model twice with the instructions and an example model available for reference before completing the model a third time without access to the instructions or example model. Participants were given the final diagram of each completed model and allowed to reference it during all recall attempts. The researcher measured the amount of time it took participants to complete the training and recall models, starting the timer when the participant
indicated they understood what was expected of them and that they were ready to begin the task, and ending the timer when the participant indicated they had completed the task. Perceived task difficulty and confidence were also collected after participants completed each training and recall model. Following each instruction condition, participants completed the perceived instruction method difficulty, workload, user experience, and trainee reaction questions. After completing all three instruction conditions, participants folded each recall model again to assess their post-task performance. Time to completion, perceived task difficulty, and confidence were collected again following completion of each post-task model. Participants were also asked to respond to the post-test open-ended questions and SSQ, concluding the first study session. Participants received their first compensation payment following the first study session.

Participants who completed the first study session and were interested in participating in the second study session were scheduled to return one week later. During the second study session, participants folded each recall model. Time to completion, perceived task difficulty, and confidence were collected following completion of each retention-test model. Participants were also asked to respond to the retention-test open-ended questions. Finally, participants were debriefed and received their second and final compensation payment. All origami models made by participants were kept in order to measure performance, which was assessed using a rubric that evaluated the accuracy of the completed models. Each participant took approximately 2.5 hours to complete the study (two hours for the first study session, 30 minutes for the second study session). Figure 12 graphically depicts the procedure sequence for each study session. Participants were compensated $10 for participating in the first session and $20 for participating in the second session.
Procedure Sequence

Study Session 1 (120 minutes)

1. Demographics & Paper Folding Test
2. Instruction Method Conditions
   - Practice Model
   - Two Training Models (with instructions)
   - One Evaluated Model (without instructions)
   - Post-Model Questionnaires
   - Time, Performance, Task Difficulty, Confidence
   - Instruction Method Difficulty, Workload, User Experience, Trainee Reactions
3. Post-Test
   - All Evaluated Models (without instructions)
   - Open-Ended Questions and SSQ
   - First payment ($10)

Study Session 2 (30 minutes)

4. Retention-Test (1 week later)
   - Time, Performance, Task Difficulty, Confidence
   - All Evaluated Models (without instructions)
   - Open-Ended Questions
   - First payment ($20)
Results

The following sections present results of the descriptive and inferential statistics performed to better understand the differences, if any, between paper-, AR-, and MR-based instructions for a procedural task. Data was analyzed using IBM SPSS Statistics 27 and Microsoft Excel.

Demographics

Thirty participants (13 male, 17 female) completed this study. Participant ages ranged from 18 to 37 years (\(Mdn = 21.5\), \(IQR = 6\)). Eight participants wore prescription glasses under the HoloLens 2 during Session 1. Three participants reported being left-handed. Twenty-three participants reported prior use of XR headsets, and eight reported owning an XR headset. Of those who used XR headsets prior to the study, 11 participants reported using AR/MR headsets for at least one hour, and 14 participants reported using VR headsets for at least one hour (see Figure 13). Five participants reported using the Microsoft HoloLens 2 prior to participating in this study.

Figure 13

Number of Hours of Reported XR Headset Use Prior to Study
Twenty-five participants reported playing video games. Table 3 shows a summary of their self-reported video game experience.

**Table 3**

*Participants’ Self-Reported Video Game Experience*

<table>
<thead>
<tr>
<th>Variable</th>
<th>n (out of 25)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of Video Game Player</strong></td>
<td></td>
</tr>
<tr>
<td>Newbie/Novice</td>
<td>3</td>
</tr>
<tr>
<td>Casual</td>
<td>15</td>
</tr>
<tr>
<td>Mid-core</td>
<td>5</td>
</tr>
<tr>
<td>Hardcore/Expert</td>
<td>2</td>
</tr>
<tr>
<td><strong>Average Hours Spent Playing Video Games Per Week</strong></td>
<td></td>
</tr>
<tr>
<td>Less than 1 hour</td>
<td>4</td>
</tr>
<tr>
<td>1 to 4 hours</td>
<td>7</td>
</tr>
<tr>
<td>5 to 9 hours</td>
<td>7</td>
</tr>
<tr>
<td>10 to 19 hours</td>
<td>5</td>
</tr>
<tr>
<td>20 to 29 hours</td>
<td>1</td>
</tr>
<tr>
<td>30 to 39 hours</td>
<td>1</td>
</tr>
<tr>
<td>More than 40 hours</td>
<td>0</td>
</tr>
<tr>
<td><strong>Devices Frequently Used to Play Video Games</strong></td>
<td></td>
</tr>
<tr>
<td>Computer Device</td>
<td>19</td>
</tr>
<tr>
<td>Console Device</td>
<td>17</td>
</tr>
<tr>
<td>Handheld Gaming Device</td>
<td>2</td>
</tr>
<tr>
<td>Mobile Device</td>
<td>13</td>
</tr>
<tr>
<td>Headset Device</td>
<td>7</td>
</tr>
<tr>
<td><strong>Frequently Played Video Game Genres</strong></td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td>18</td>
</tr>
<tr>
<td>Adventure</td>
<td>14</td>
</tr>
<tr>
<td>Driving</td>
<td>11</td>
</tr>
<tr>
<td>Educational/Edutainment</td>
<td>0</td>
</tr>
<tr>
<td>Fighting</td>
<td>5</td>
</tr>
<tr>
<td>Fitness</td>
<td>2</td>
</tr>
<tr>
<td>Music/Dance</td>
<td>2</td>
</tr>
<tr>
<td>Puzzle/Card</td>
<td>7</td>
</tr>
<tr>
<td>Retro/Classic</td>
<td>3</td>
</tr>
<tr>
<td>Role Playing</td>
<td>7</td>
</tr>
<tr>
<td>Simulation</td>
<td>11</td>
</tr>
<tr>
<td>Social/Social Network</td>
<td>3</td>
</tr>
<tr>
<td>Sports</td>
<td>3</td>
</tr>
<tr>
<td>Strategy</td>
<td>8</td>
</tr>
</tbody>
</table>

*Note. *Indicates questions in which participants were permitted to select multiple response choices.*
In general, participants did not report much familiarity or proficiency with origami (see Figures 14-15).

**Figure 14**

*Self-Reported Origami Familiarity*

![Origami Familiarity graph](image)

**Figure 15**

*Self-Reported Origami Proficiency*

![Origami Proficiency graph](image)
Seventeen participants returned for their second session seven days after completing their first session. Other participants returned for their second session less than \((n = 1)\) or more than seven days \((n = 12)\) after completing their first session due to scheduling difficulties caused by two hurricanes that impacted the Daytona Beach area and disrupted campus operations during the Fall 2022 semester (see Table 4).

**Table 4**

*Number of Days Between Participants’ First and Second Study Session*

<table>
<thead>
<tr>
<th>Number of Days Between First and Second Session</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
</tr>
</tbody>
</table>

**Paper Folding Test**

Twelve participants attempted all twenty questions of the paper folding test. Across all participants, the average number of correct responses was 14.13 \((SE = 0.61)\). When considering the proportion of correct to attempted answers, participants’ average score was 78.3\% \((SE = 3.1\%)\). The average score out of all twenty questions was 70.7\% \((SE = 3.1\%)\). These results are comparable to those reported in the cognitive test kit from which the paper folding test originates, where the average number of correct responses collected from 46 college students was 13.8 \((SD = 4.5; Ekstrom et al., 1976)\). The current study’s results are also comparable to a more recent sample of 149 participants aged 19-64 years, where the average number of correct responses was 12.7 \((SD = 3.5; Burte et al., 2018)\).
Summary of Measures and Inferential Statistics Assumptions

Table 5 presents a summary of measures collected across each condition of the study.

Table 5

Measures for Each Study Condition

<table>
<thead>
<tr>
<th>Instruction Method</th>
<th>Time of Procedure Recall</th>
<th>Immediate</th>
<th>Post-Test</th>
<th>Retention-Test (1 week later)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper</td>
<td>Conditions 1-3</td>
<td>Performance</td>
<td>Time to Completion</td>
<td>Task Difficulty</td>
</tr>
<tr>
<td></td>
<td>Instruction Method</td>
<td>Confidence</td>
<td>Workload</td>
<td>User Experience</td>
</tr>
<tr>
<td>AR (images only)</td>
<td></td>
<td>Trainee Reactions</td>
<td>Cybersickness</td>
<td></td>
</tr>
<tr>
<td>MR (images and</td>
<td>Conditions 4-9</td>
<td>Performance</td>
<td>Time to Completion</td>
<td>Task Difficulty</td>
</tr>
<tr>
<td>virtual cues)</td>
<td></td>
<td>Confidence</td>
<td>Workload</td>
<td>User Experience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Open-Ended Questions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ANOVAs are appropriate analyses for data that meets the following assumptions: 1) dependent variable is continuous; 2) Independent variable has three or more levels; 3) there are no outliers in the data set; 4) the sampling distribution of means for each level of the independent variable(s) is normal; 5) homogeneity of variance across all levels of the independent variable (Tabachnick & Fidell, 2006). The experimental design of this study satisfied Assumptions 1 and 2. Determining satisfaction of Assumption 3 was completed by identifying data points greater than ±3 standard deviations from the mean (Goodwin & Goodwin, 2017). Twenty-four outliers were identified across all of the dependent variables, which are specified in Appendix J. Unless otherwise stated, outliers are included in the following results sections because including the outliers did not substantially impact the interpretation of the ANOVA analyses. Regarding Assumption 4, normality was assessed using skewness and kurtosis. Satisfaction of
Assumption 5 was determined using Mauchly’s test of sphericity. Results in which Mauchly’s test of sphericity was significant, indicating violation of Assumption 5, are reported using the Greenhouse-Geisser correction.

**Performance**

To measure performance, three raters used scoring rubrics to evaluate the accuracy of participants’ Training 2, Immediate, Post-Test, and Retention attempts for the Cicada, Samurai Helmet, and Necktie. Percent agreement was calculated by summing the number of instances each pair of raters came to complete agreement and dividing this sum by the total number of rated models. Overall percent agreement was 75.40% for the Cicada, 84.05% for the Samurai Helmet, and 84.95% for the Necktie. Cronbach’s alpha can be used as a measure of inter-rater reliability to determine the extent to which raters generate corresponding scores (DeVellis, 2003). Cronbach’s alpha coefficients greater than .70 are considered acceptable (Tavakol & Dennick, 2011). Cronbach’s alpha was .924 for the Cicada, .946 for the Samurai Helmet, and .964 for the Necktie, indicating acceptable correspondence between rater scores for all models.

A two-way repeated measures ANOVA was conducted to determine the effect of instruction method and time of procedure recall on performance. Overall performance scores can range from 0 to 7, with higher scores indicating more accurate models. The main effect of recall time showed a significant difference in performance across recall time periods, $F(1.47, 42.71) = 25.05, p < .005$, partial $\eta^2 = .46$. Retention models ($M = 3.07, SE = 0.37$) were significantly less accurate than the Immediate ($M = 4.97, SE = 0.15$) and Post models ($M = 4.25, SE = 0.26$). Post models were also significantly less accurate than the Immediate models (see Figure 16). There was no statistically significant main effect of instruction method, $F(2, 58) = 0.63, p = .538$, or interaction between instruction method and time of procedure recall, $F(4, 116) = 1.24, p = .297$. Hypothesis 1 was not supported, while Hypothesis 2 was partially supported.
Comparison of Performance Between Instruction Methods Over Time

**Figure 16**

*Note.* Error bars represent ±1 standard error.

**Time to Completion**

A two-way repeated measures ANOVA was conducted to determine the effect of instruction method and time of procedure recall on completion time, which was measured in seconds. An assessment of studentized residuals greater than ±3 standard deviations identified three outliers (see Appendix J). Prior to completing the ANOVA, the three outliers were replaced with the average time for all other completion times of that condition. The main effect of time of procedure recall showed a significant difference in completion time across recall time periods, $F(1.52, 43.93) = 4.39$, $p = .027$, partial $\eta^2 = .13$. Post-hoc analyses did not find significant differences between Immediate ($M = 147.03$, $SE = 8.67$), Post-Test ($M = 149.24$, $SE = 7.95$), or Retention ($M = 174.49$, $SE = 12.12$) completion times (see Figure 17). Hypothesis 4 was not supported. There was no statistically significant main effect of
instruction method, $F(2, 58) = 0.07, p = .934$, or interaction between instruction method and time of procedure recall, $F(4, 116) = 0.94, p = .442$. Hypothesis 3 was not supported.

**Figure 17**

*Comparison of Completion Time Between Instruction Methods Over Time*

![Graph showing completion time comparison](image)

*Note.* Error bars represent ±1 standard error.

Regarding Exploratory Hypothesis 14, a two-way repeated measures ANOVA was conducted to investigate differences in completion time between the two training sessions. The main effect of recall time showed a significant difference in completion time between the two training sessions $F(1, 29) = 36.41, p < .005$, partial $\eta^2 = .56$. Training 1 ($M = 220.62, SE = 11.64$) took significantly longer to complete than Training 2 ($M = 180.51, SE = 10.36$). There was no statistically significant main effect of instruction method, $F(2, 58) = 1.53, p = .226$, or interaction between instruction method and time of procedure recall, $F(2, 58) = 0.58, p = .561$. 

60
**Perceived Difficulty**

A two-way repeated measures ANOVA was conducted to determine the effect of instruction method and time of procedure recall on perceived instruction method difficulty. Perceived instruction method difficulty was self-reported by participants using a 7-point scale (1 - Very Difficult to 7 - Very Easy) after completing each instruction method during Session 1. Shown in Figure 18, there was no statistically significant difference of perceived difficulty ratings between instruction method, \( F(2, 58) = 0.76, p = .473 \). Hypothesis 5 was not supported.

**Figure 18**

*Comparison of Perceived Instruction Method Difficulty*

![Comparison of Perceived Instruction Method Difficulty](image)

*Note.* Error bars represent ±1 standard error. 1 = Very Difficult; 7 = Very Easy.

A two-way repeated measures ANOVA was conducted to determine the effect of instruction method and time of procedure recall on perceived task difficulty. Perceived task difficulty was self-
reported by participants using a 7-point scale (1 - Very Difficult to 7 - Very Easy) after each model attempt. The main effect of recall time showed a significant difference in perceived task difficulty across recall time periods, $F(1.59, 45.97) = 31.55, p < .005$, partial $\eta^2 = .52$. Retention models ($M = 3.52, SE = 0.30$) were perceived to be significantly more difficult than Immediate ($M = 5.41, SE = 0.16$) and Post-Test models ($M = 4.66, SE = 0.23$). Post-test models were perceived to be significantly more difficult than Immediate models (see Figure 19). Hypothesis 6 was supported. There was no statistically significant main effect of instruction method, $F(2, 58) = 0.07, p = .932$, or interaction between instruction method and time of procedure recall, $F(4, 116) = 0.76, p = .555$.

**Figure 19**

*Comparison of Perceived Task Difficulty Between Instruction Methods Over Time*

![Graph showing comparison of perceived task difficulty between instruction methods over time.](image)

*Note.* Error bars represent ±1 standard error. 1 = Very Difficult; 7 = Very Easy.

Regarding Exploratory Hypothesis 14, a two-way repeated measures ANOVA was conducted to investigate differences in perceived task difficulty between the two training sessions. The main effect of
recall time showed a significant difference in perceived task difficulty between the two training sessions, $F(1, 29) = 70.44, p < .005$, partial $\eta^2 = .71$. Training 1 models ($M = 4.31$, $SE = 0.19$) were perceived to be significantly more difficult than the Training 2 models ($M = 5.34$, $SE = 0.16$). There was no statistically significant main effect of instruction method, $F(2, 58) = 0.50, p = .610$, or interaction between instruction method and time of procedure recall, $F(2, 58) = 0.85, p = .434$.

A one-way repeated measures ANOVA was conducted to determine whether perceived difficulty scores differed across models. There was a significant difference in difficulty ratings between models, $F(2, 298) = 5.07, p = .007$, partial $\eta^2 = .03$. Helmet ($M = 4.92$, $SE = 0.16$) was rated significantly easier than Necktie ($M = 4.47$, $SE = 0.16$; see Figure 20).

**Figure 20**

*Comparison of Perceived Difficulty Between Origami Models*

![Bar chart comparing perceived difficulty of Cicada, Helmet, and Necktie models.](image)

*Note.* Error bars represent $\pm 1$ standard error.
Perceived Confidence

A two-way repeated measures ANOVA was conducted to determine the effect of instruction method and time of procedure recall on perceived confidence. Perceived confidence was self-reported by participants using a 7-point scale (1 - Not at all Confident to 7 - Very Confident) after each model attempt. The main effect of recall time showed a significant difference in perceived confidence across recall time periods, $F(1.61, 46.71) = 33.40, p < .005$, partial $\eta^2 = .54$. Participants were significantly less confident about their Retention model success ($M = 3.60, SE = 0.35$) compared to their Immediate ($M = 5.71, SE = 0.20$) and Post-Test Models ($M = 5.00, SE = 0.28$). Participants were also significantly less confident about their Post-Test models compared to their Immediate models (see Figure 21). Hypothesis 8 was supported. There was no statistically significant main effect of instruction method, $F(1.64, 47.42) = 0.19, p = .782$, or interaction between instruction method and time of procedure recall, $F(4, 116) = 1.10, p = .358$. Hypothesis 7 was not supported.

Regarding Exploratory Hypothesis 14, a two-way repeated measures ANOVA was conducted to investigate differences in perceived confidence between the two training sessions. The main effect of recall time showed a significant difference in perceived confidence between the two training sessions, $F(1, 29) = 13.35, p = .001$, partial $\eta^2 = .32$. Participants were significantly less confident about their Training 1 model success ($M = 5.20, SE = 0.24$) compared to their Training 2 model ($M = 5.78, SE = 0.20$). There was no statistically significant main effect of instruction method, $F(1.54, 44.69) = 0.62, p = .503$, or interaction between instruction method and time of procedure recall, $F(2, 58) = 1.06, p = .352$. 
A one-way repeated measures ANOVA was conducted to determine whether perceived confidence scores differed across models. There was a significant difference in confidence ratings between models, $F(2, 298) = 14.90, p < .005$, partial $\eta^2 = .05$. Helmet ($M = 4.56, SE = 0.15$) was rated significantly higher than Necktie ($M = 4.47, SE = 0.14$; see Figure 22).
Comparison of Perceived Confidence Between Origami Models

Note. Error bars represent ±1 standard error. 1 = Not at all Confident; 7 = Very Confident.

Workload

A series of one-way repeated measures ANOVAs was conducted to determine whether raw ratings of each NASA-TLX dimension differed between instruction methods. Shown in Figure 23, there were no statistically significant differences in workload between instruction methods: Mental, $F(1.44, 41.68) = 1.64, p = .210$; Physical, $F(2, 58) = 0.07, p = .932$; Temporal, $F(2, 58) = 3.15, p = .050$, partial $\eta^2 = .10$; Performance, $F(2, 58) = 0.27, p = .765$; Effort, $F(2, 58) = 3.00, p = .057$, partial $\eta^2 = .09$; and Frustration, $F(1.63, 47.18) = 3.03, p = .068$, partial $\eta^2 = .10$. Hypothesis 9 was not supported.
**Comparison of Workload Ratings Between Instruction Methods**

![Comparison of Workload Ratings Between Instruction Methods](image)

*Note.* Error bars represent ±1 standard error. Higher ratings indicate that participants perceived the task as more demanding or that they performed poorly.

**User Experience**

User experience was measured using the UEQ-S, an 8-item questionnaire that was distributed to participants following their exposure to each of the three instruction methods. The UEQ-S was analyzed by averaging ratings for the first four items to produce a pragmatic quality score, averaging the last four items to produce a hedonic quality score, and averaging all eight items to produce an overall user experience score. Benchmarks for each score (Excellent, Good, Above Average, Below Average, and Bad) were calculated using the analysis tool provided by the UEQ developers (Hinderks et al., 2018).

A series of one-way repeated measures ANOVAs was conducted to determine whether pragmatic quality, hedonic quality, and overall scores differed across instruction methods. There was no statistically significant difference in average pragmatic quality scores between instruction methods, $F(2, 58) = 0.54, p = .587$. Hypothesis 10 was not supported. Benchmarks for pragmatic quality were Good for
Paper, Above Average for AR, and Good for MR. There was a statistically significant difference in average hedonic quality scores between instruction methods, $F(1.61, 46.64) = 70.60, p < .005$, partial $\eta^2 = .71$. MR scores ($M = 6.23, SE = 0.12$) were significantly higher than AR scores ($M = 5.70, SE = 0.20$), which were significantly higher than Paper scores ($M = 3.57, SE = 0.24$). Hypothesis 11 was supported.

Benchmarks for hedonic quality were Bad for Paper, Excellent for AR, and Excellent for MR. There was a statistically significant difference in average overall user experience scores between instruction methods, $F(2, 58) = 20.54, p < .005$, partial $\eta^2 = .42$. Paper scores ($M = 4.62, SE = 0.17$) were significantly lower than AR ($M = 5.55, SE = 0.17$) and MR scores ($M = 5.91, SE = 0.13$; see Figure 24). There was no statistically significant difference between AR and MR scores. Benchmarks for overall scores were Below Average for Paper, Good for AR, and Excellent for MR.

**Figure 24**

*Comparison of User Experience Ratings Between Instruction Methods*

![Comparison of User Experience Ratings Between Instruction Methods](image)

*Note.* Error bars represent ±1 standard error.
Trainee Reactions

Trainee reactions were collected using a six-item questionnaire adapted from Long et al. (2008) after participants’ exposure to each instruction method. Participants rated each item on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree). Item numbers referenced in the following paragraphs correspond to item numbers listed in Table 6.

Table 6

Trainee Reactions Questionnaire Items

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Trainee Reaction Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The training content was clear.</td>
</tr>
<tr>
<td>2</td>
<td>I could easily understand the training content.</td>
</tr>
<tr>
<td>3</td>
<td>I was able to navigate through the training content.</td>
</tr>
<tr>
<td>4</td>
<td>I found the training content easy to use.</td>
</tr>
<tr>
<td>5</td>
<td>I was satisfied with the presentation of the training content.</td>
</tr>
<tr>
<td>6</td>
<td>I had a positive learning experience.</td>
</tr>
</tbody>
</table>

A series of one-way repeated measures ANOVAs was conducted to determine whether the six individual trainee reaction items differed between instruction methods. There were significant differences between instruction methods for Item 3, \( F(2, 58) = 6.39, p < .005, \text{ partial } \eta^2 = .18 \), and Item 4, \( F(1.53, 44.21) = 5.49, p = .013, \text{ partial } \eta^2 = .16 \). Paper (\( M = 4.97, SE = .03 \)) was rated significantly higher than AR (\( M = 4.53, SE = .12 \)) and MR (\( M = 4.57, SE = .10 \)) for Item 3, indicating that participants were better able to navigate through training content using the paper instructions. Paper (\( M = 4.83, SE = .07 \)) was also rated significantly higher than AR (\( M = 4.37, SE = .18 \)) and MR (\( M = 4.43, SE = .09 \)) for Item 4, indicating that it was easier for participants to use the paper instructions. Hypothesis 12 was partially supported. There was no significant difference in ratings for Item 1, \( F(2, 58) = .053, p = .590 \); Item 2, \( F(1.62, 46.99) = 0.79, p = .437 \); Item 5, \( F(1.53, 44.27) = 2.56, p = .086, \text{ partial } \eta^2 = .08 \); or Item 6, \( F(2, 58) = 0.03, p = .969 \). See Figure 25 for a comparison of trainee reaction item responses across instruction methods.
**Figure 25**

*Comparison of Trainee Reaction Ratings Between Instruction Methods*

![Graph showing comparison of reaction ratings between different instruction methods.](image)

*Note.* Error bars represent ±1 standard error. Item 1 = The training content was clear. Item 2 = I could easily understand the training content. Item 3 = I was able to navigate through the training content. Item 4 = I found the training content easy to use. Item 5 = I was satisfied with the presentation of the training content. Item 6 = I had a positive learning experience.

Participants were also asked to answer three open-ended response questions regarding their likes, dislikes, and recommendations for improvement for each instruction method. These findings are associated with Exploratory Hypothesis 16 and outlined in Tables 7-8.
### Table 7

**Participant Likes for Each Instruction Method**

<table>
<thead>
<tr>
<th>Instruction Method</th>
<th>Likes</th>
<th>n</th>
<th>Representative Participant Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper AR MR</td>
<td>X Presence of virtual cues</td>
<td>21</td>
<td>“Being able to see the folding lines was extremely useful in learning where to accurately perform each fold.”</td>
</tr>
<tr>
<td></td>
<td>X Easy to navigate and use</td>
<td>13</td>
<td>“I liked that I didn’t have to wait and hold down on a button before I could go to the next step. I could just fly through the pages at my own pace.”</td>
</tr>
<tr>
<td></td>
<td>X X Voice commands and ability to use hands-free</td>
<td>5</td>
<td>“AR training was cool. It was nice to be able to say ‘next step’ and it would automatically move through, and I didn’t have to use my hands.”</td>
</tr>
<tr>
<td></td>
<td>X Absence of virtual cues</td>
<td>3</td>
<td>“I liked how it was very easy to look at the instructions and then fold without the mixed reality arrows getting in the way.”</td>
</tr>
</tbody>
</table>
Table 8

**Participant Dislikes for Each Instruction Method**

<table>
<thead>
<tr>
<th>Instruction Method</th>
<th>Dislikes</th>
<th>n</th>
<th>Representative Participant Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper</td>
<td>X Virtual cues were obstructive and distracting</td>
<td>21</td>
<td>“The overlaid graphics sometimes made it more difficult to see my hands and what I was doing.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“The instructions on the grid were a little distracting and off-centered.”</td>
</tr>
<tr>
<td>AR</td>
<td>X Navigation and hardware issues</td>
<td>19</td>
<td>“The navigation controls were a bit laggy at times, causing me to go back and forward through the steps unintentionally.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“The HoloLens is slightly tinted, darkening the real-world elements significantly and making lining up the paper edges more difficult.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“The screen on the headset is short.” [Referencing limited field of view; FOV]</td>
</tr>
<tr>
<td>MR</td>
<td>X Required more head movement</td>
<td>7</td>
<td>“Looking back and forth between the AR display and the real-world object was a bit disorienting.”</td>
</tr>
<tr>
<td></td>
<td>X Having to manually flip the pages</td>
<td>4</td>
<td>“I had to take my hands off the [origami] paper when flipping the [instruction] page, so the [origami] paper would frequently move from its position on the grid.”</td>
</tr>
<tr>
<td></td>
<td>X Not as comprehensive as AR/MR instructions</td>
<td>3</td>
<td>“It was easy to follow [the paper instructions], but I felt like I was missing instructions compared to the AR/MR.”</td>
</tr>
</tbody>
</table>
To summarize participants’ likes and dislikes, participants found the paper instructions to be familiar and reported paper easier to navigate and use compared to the AR/MR instructions. However, participants did not enjoy having to take their hands off their origami paper in order to use the paper instructions, and liked the voice commands and eye gaze features offered by the AR/MR training that enabled them to interact with the instructions hands-free. Several participants shared that the MR virtual cues (e.g., dotted lines to indicate fold placement, arrows to indicate fold direction) were the best aspect of the MR training because they promoted accuracy and reduced the need to move their head to view the training materials, but some also perceived the MR virtual cues as distracting and obstructive. Participants also noted more navigation issues with the AR/MR training compared to the paper instructions, primarily referencing issues with lag. Other AR/MR dislikes included limited FOV, increased head movement to reference AR instructions, and the visor tint that negatively impacted participants’ ability to see their real-world surroundings.

When asked how they would improve the paper training, thirteen participants suggested adding more written (e.g., “Adding words to the instructions on how and where exactly I’m supposed to make certain folds and creases would be beneficial.”) or visual instructions (e.g., “More pictures indicating different angles,” and, “Put a small image of what it should look like when you are done with the step up in the corner.”) to the paper training. Three participants stated they would put all of the paper instructions on one page to reduce the need to flip between steps (e.g., “It would be easier to have all the steps on one sheet, that way you can see how each step leads to the next.”). Regarding MR training improvements, twelve participants provided suggestions for redesigning the virtual content to make it less obstructive, such as making the virtual cues thinner or more transparent, implementing the ability to customize the color of the virtual cues, reducing the number of virtual cues by only showing the dotted fold lines and removing the arrows, and toggling the presence and absence of the virtual cues depending on whether the user’s hands are in the workspace (e.g., “Have the arrows and words that
pop up on the grid disappear when your hands are there, but reappear when you move your hands away, so you can reference the instructions at any time just by staring at the grid.”). Five participants suggested adding folding animations or auditory instructions to the AR/MR training conditions. Four participants recommended better alignment of the MR virtual cues to the real world.

Cybersickness

Cybersickness was measured using the SSQ, a sixteen-item questionnaire that was distributed to participants at the end of their first study session. The following benchmarks can be used to facilitate interpretation of SSQ ratings: 0 – No Symptoms; <5 - Negligible Symptoms; 5-10 - Minimal Symptoms; 10-15 - Significant Symptoms; 15-20 - Concerning Symptoms; and >20 - Bad (Stanney et al., 1997). Oculomotor discomfort ($M = 17.18, SD = 15.28$) was the highest score among the subscales and indicated Concerning Symptoms. Disorientation ($M = 14.85, SD = 14.13$) and total SSQ scores ($M = 10.60, SD = 8.68$) were Significant, while nausea ($M = 6.04, SD = 7.72$) was Minimal.

Relationships Between Demographic Variables and Performance

Per Exploratory Hypothesis 13, Pearson’s correlations were run to investigate the relationships between select demographic variables (e.g., spatial visualization ability; prior video game and origami experience) and performance. Spatial visualization ability was measured using the paper folding test (PFT). As shown in Table 9, there were statistically significant, moderate positive correlations between PFT scores and paper performance ($r(28) = .60, p < .005$), between PFT scores and AR performance ($r(28) = .45, p = .012$), and between PFT scores and MR performance ($r(28) = .43, p = .018$). These results indicate that as spatial visualization ability increased, performance for all instruction methods also increased. Regarding origami experience, there were statistically significant, moderate positive correlations between origami experience and paper performance ($r(28) = .39, p = .031$) and between origami experience and AR performance ($r(28) = .40, p = .030$). There were no significant correlations between origami experience and MR performance, or between performance and average number of
hours playing video games per week. A t-test comparing performance between those with and without self-reported prior XR usage found no significant difference in paper, AR, or MR performance.

Table 9

*Correlation Matrix for Select Demographic Variables and Performance*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>.42*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR</td>
<td>.18</td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PFT</td>
<td>.60**</td>
<td>.45*</td>
<td>.43*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video Game Usage</td>
<td>-.07</td>
<td>-.12</td>
<td>-.001</td>
<td>-.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origami Experience</td>
<td>.39*</td>
<td>.40*</td>
<td>.19</td>
<td>.39*</td>
<td>-.17</td>
<td></td>
</tr>
</tbody>
</table>

*Note. *Correlation significant at the 0.05 level. **Correlation significant at the 0.01 level.*

Post- and Retention-Test Open-Ended Questions

Following the Session 1 post-test and Session 2 retention-test of all three origami models, participants were asked to rank their preference for instructions provided using paper, AR, or MR, from most to least preferred. Additionally, after the retention-test, participants were asked to describe the strategies they used to recall each model and how each instruction method affected their ability to recall the procedure. These findings are associated with Exploratory Hypotheses 15 and 16 and presented in the following sections.

Preference

At the end of the first and second sessions, participants were asked to rank paper, AR, and MR from their most (1) to least (3) preferred instruction method. A Friedman test was conducted to determine if there were differences in how instruction methods were ranked. There was a statistically significant difference between average ranks of the instruction methods collected at the end of the first study session, \( \chi^2(2) = 6.07, p = .048 \). MR was ranked first by 15 participants \( (M = 1.80) \), paper was ranked
first by 9 participants ($M = 1.83$), and AR was ranked first by 6 participants ($M = 2.37$). Post-hoc analysis with Wilcoxon signed-rank tests conducted with a Bonferroni correction found a significant increase in average rank for paper compared to AR ($p = .021$). There was no significant difference between AR and MR ($p = .063$) or paper and MR ($p = .922$) average rankings. There was no significant difference between average ranks collected at the end of the second study session, $\chi^2(2) = 5.40, p = .067$. MR was ranked first by 16 participants ($M = 1.70$), paper was ranked first by 8 participants ($M = 2.00$), and AR was ranked first by 6 participants ($M = 2.30$; see Figure 26). These results suggest that AR is the least preferred method of instruction.

**Figure 26**

*Comparison of Preference Between Instruction Methods*

![Bar chart showing preference rankings between instruction methods for two sessions.](image)

*Note.* Error bars represent ±1 standard error. Lower average rankings indicate a more preferred instruction method.

In addition to providing their rankings for each instruction method, participants were asked to explain why they ranked the instruction methods in the order they chose.
**Comments from Participants Who Preferred Paper.** Participants who ranked paper instructions as their most preferred instruction method explained they were more familiar with paper instructions and felt the paper instructions were less obtrusive and effortful. For example, one participant commented on their familiarity with paper instructions by stating, “Using paper instructions is just a habit, it is what I work with all the time. It is like having a textbook and homework laid out on a desk. I am used to that setup,” while another said, “The MR was helpful, interesting, and different, but if I was learning something for the first time, I would want to stick with something I was used to using, like paper.” Perceptions of AR/MR obtrusiveness primarily stemmed from the MR virtual cues, “MR was obstructive. The 3D content obstructed my view of the paper, the grid, and my hands. I could not see what I was folding.” The headset hardware also contributed to its obstructiveness, “Even when I adjusted the headset, it was a little blurry to look through it. Paper is very sharp and clean.” Paper was also considered to be less effortful to navigate through the training materials, “I can flip through it quickly or close it altogether if I don’t want it. But with AR and MR, I have to stare at the button for a few seconds to flip the page. I can’t flip multiple pages or flip very quickly in AR and MR because I have to wait for it to respond to my input.” Several participants who ranked paper as their most preferred method noted that AR presented the training materials in a very similar manner, but with the added difficulty of having to learn and use a new device and application (e.g., “AR seemed unnecessary for this task and the headset started to weigh on my neck a little bit. I didn’t feel like it added anything I couldn’t have just gotten from the paper,” and, “Paper and AR are more or less the same thing, but AR involves wearing a headset. Why wear a headset when you don’t have to?”).

**Comments from Participants Who Preferred AR.** Participants who ranked AR instructions as their most preferred instruction method stated they found AR more exciting, enjoyed the ability to navigate through the training materials hands-free, and perceived it to be less obstructive. For those who preferred AR, AR was considered more exciting than paper, “The paper instructions were easy to
follow. I just preferred the AR more because it was interactive, entertaining, and interesting.” Other participants shared, “I like how innovative the AR is and I like the fact that I don’t have to remove my hands from the origami to look at the instructions,” and, “The AR allowed me to fold better. When I had to flip the paper pages, the model would move since I had to take my hands off of it. It was inconvenient to have to take my hands off the model.” The perceived obtrusiveness of the MR virtual cues was exacerbated by the HoloLens 2 limited FOV, “AR was the easiest to work with. You could look back and forth, but you didn’t have to look down at the book because it was just in front of you. MR was difficult to get used to. It was weird having to look down at it because it [the virtual cues] wouldn’t show up if you just glanced down, so you had to make sure to really look at it. And when I was folding, it would get in the way so I couldn’t see if my folds were exactly straight or not.”

Comments from Participants Who Preferred MR. Participants who most preferred MR instructions indicated they found MR more integrative and comprehensive, providing benefits that outweigh the disadvantages of using a headset. Several participants noted that because the MR instructions were better integrated with the participants’ workspace, they experienced less cognitive workload and reduced head movement. For instance, one participant stated, “With MR, I don’t have to interact with so many other things around me. I just look at the [origami] paper and fold it as I am looking at it, instead of having to look up at the AR and take time to look away from my paper to find the screen, then reorient myself to what I was doing with the paper,” and another participant commented, “Because the MR placed the instructions right on my paper, I didn’t have to hold the information in my head as long or keep looking back at the instructions, as I did with the paper and AR instructions.” Additionally, those who preferred MR training believed it provided more information than the other instruction methods, “I felt like the MR provided the most in-depth instruction and it was easiest to follow along because it was so immersive.” Other participant comments that support this notion include, “I liked that MR provided multiple sources of information. I was able to be more precise with
my folds because I had so many cues to rely on,” and, “MR was a lot more helpful. It took the guesswork out of it. The dotted lines and arrows really showed exactly where to fold.” Participants who ranked MR first also noted that AR did not provide additional benefits that outweigh the disadvantages of using the headset. Representative participant comments for this point include, “I didn’t feel like the AR contributed anything more than paper, but AR made it more cumbersome to complete the task,” and, “For the AR, I feel like it wasn’t providing any value. It was like using technology for technology’s sake, because it was almost exactly what the paper offered.”

**Recall Strategies**

At the end of the second study session, participants were asked to reflect upon and share the strategies they used to recall the procedure for each origami model. Twenty-four participants stated they generated and utilized mental images as they recalled the models. Table 10 provides a summary of these visualization strategies.
### Table 10

*Visualization Strategies Used by Participants to Recall the Retention Models*

<table>
<thead>
<tr>
<th>Instruction Method</th>
<th>Visualization Strategy</th>
<th>n</th>
<th>Representative Participant Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper X AR X MR X</td>
<td>Pictured the instructions</td>
<td>9</td>
<td>“I was trying to remember the pictures [step-by-step diagrams] of the folds.”</td>
</tr>
<tr>
<td>X</td>
<td>Imagined the process of completing each model</td>
<td>9</td>
<td>“I was visualizing myself folding the paper in my head.”</td>
</tr>
<tr>
<td>V</td>
<td>Visualized the MR virtual cues</td>
<td>8</td>
<td>“I was picturing the dotted lines and arrows, which was helpful.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“The MR was so distracting that I could not think of the steps, only the arrows.”</td>
</tr>
<tr>
<td>X</td>
<td>Pictured an evolving model</td>
<td>6</td>
<td>“I was anticipating what the paper should look like and then folding it until it matched that image.”</td>
</tr>
</tbody>
</table>

*Note. n out of the 24 participants who stated they generated and utilized mental images as they recalled the models. Participants from this sample subset may have indicated multiple visualization strategies.*

As shown in Table 10, eight participants indicated they visualized the MR virtual cues. Of these eight participants, six commented that this visualization strategy facilitated their ability to recall the models they learned using MR (e.g., “I was picturing the dotted lines and arrows, which was helpful.”), one said they did not think it helped their recall ability, and one stated it hindered their ability to accurately recall the model (e.g., “The MR was so distracting that I could not think of the steps, only the arrows.”).

Other recall strategies mentioned by participants include relying on muscle memory, leveraging the diagram of the finished model provided to participants during all recall models, and using points on the grid to help them determine where to make folds. These strategies are summarized in Table 11.
Table 11

*Other Strategies Used by Participants to Recall the Retention Models*

<table>
<thead>
<tr>
<th>Instruction Method</th>
<th>Strategy</th>
<th>n</th>
<th>Representative Participant Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper</td>
<td>Utilized muscle memory</td>
<td>18</td>
<td>“Muscle memory definitely played a role in completing certain steps, like</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>flipping the paper over.”</td>
</tr>
<tr>
<td></td>
<td>Leveraged diagram of finished model</td>
<td>18</td>
<td>“I used the diagram [of the finished model] to see if I could recall a starting</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>point or determine what the next step would be.”</td>
</tr>
<tr>
<td></td>
<td>Used points on the grid</td>
<td>10</td>
<td>“A couple of times, I lined it [the paper] up with the grid to see if it would trigger</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a memory.”</td>
</tr>
</tbody>
</table>

Note. *n* out of all 30 participants.

In general, recall strategies did not differ between instruction methods, with the exception of visualizing the virtual cues that were only available during MR training.

**Perceptions on Whether Instruction Method Matters**

At the end of the second study session, participants were also asked whether they believed learning the origami models using different methods of instruction impacted their ability to recall each model. Twenty-nine participants agreed that the instruction method influenced their recall performance, and one participant stated, “I don’t [agree] because I forgot them all”. Six participants stated the AR/MR training experiences were more fun and exciting. Of these six participants, three believed this positively contributed to their ability to recall the origami models (e.g., “I think I remembered the AR and MR models better because it was more exciting to use something other than the book.”), while two participants thought the fun experience negatively contributed to their recall performance (e.g., “I was more focused on trying to figure out how to use the headset. It was cool, but
with the paper instructions, there was nothing else to focus on but the instructions and I think I learned better as a result.”).

Six participants also mentioned the MR training was more helpful than the paper and AR training, enabling them to better remember the MR models. For example, one participant supported this notion by stating, “By placing the instructions directly on the [origami] paper, the MR provided more landmarks than the AR and paper instructions. I would argue this helped me remember the MR model better,” while another participant explained, “The MR instructions provided more guidance, which helped me learn the model better and led to better [memory] encoding.” However, three participants stated they became reliant on the MR virtual cues, which negatively impacted their recall performance. For example, one participant shared, “There were some steps that I completely missed because I was relying on the MR but did not have the lines on the [origami] paper anymore. I was definitely relying on the MR more than the other two [instruction] methods,” while another participant said, “I wonder if I relied more on the MR rather than actually learning from it. I felt a lot more lost once the MR virtual elements were gone compared to just losing the instructions with the other two methods.”

Overall, almost all of the participants believed their recall performance in the second study session was impacted by how they learned each origami model. More participants commented that they liked the AR/MR training, but they had differing opinions as to whether it helped or hindered their ability to recall each model.

**Results Summary**

Analysis of thirty participants revealed decreased performance, increased perceived difficulty, and decreased confidence as time since training increased, with no significant difference in these measures between instruction methods. Completion times and workload were found to be comparable between instruction methods. Measures with differences between instruction methods included user
experience, with higher ratings for MR, and preference, with lower rankings for AR. See Table 12 for a summary of the hypotheses and outcomes of the performed analyses.

**Table 12**

*Summary of Hypotheses and Outcomes of Performed Analyses*

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H1.</strong> Performance of the paper folding procedural task will be influenced by instruction method in that paper and MR will result in higher accuracy than AR.</td>
<td>Hypothesis not supported; There was no statistically significant main effect of instruction method.</td>
</tr>
<tr>
<td><strong>H2.</strong> Performance of the paper folding procedural task will be influenced by time of recall in that immediate recall will result in higher accuracy than post-test, followed by retention-test. The difference in performance between immediate recall and retention-test will be greater for the AR condition.</td>
<td>Hypothesis partially supported; Immediate &gt; Post &gt; Retention; No significant interaction effect.</td>
</tr>
<tr>
<td><strong>H3.</strong> Time to complete the paper folding procedural task will be influenced by instruction method in that paper and MR will result in faster completion times than AR.</td>
<td>Hypothesis not supported; There was no statistically significant main effect of instruction method.</td>
</tr>
<tr>
<td><strong>H4.</strong> Time to complete the paper folding procedural task will be influenced by time of recall in that retention-test will take more time to complete than post-test, followed by immediate recall.</td>
<td>Hypothesis not supported; There was no statistically significant difference between immediate, post, or retention.</td>
</tr>
<tr>
<td><strong>H5.</strong> Perceived difficulty of the training experience will be influenced by instruction method in that AR and MR will result in higher perceived difficulty than paper.</td>
<td>Hypothesis not supported; There was no statistically significant main effect of instruction method.</td>
</tr>
<tr>
<td><strong>H6.</strong> Perceived difficulty of the paper folding procedural task will be influenced by time of recall in that retention-test will result in higher difficulty than post-test, followed by immediate recall.</td>
<td>Hypothesis supported; Retention &gt; Post &gt; Immediate</td>
</tr>
<tr>
<td><strong>H7.</strong> Confidence associated with the paper folding procedural task will be influenced by instruction method in that AR and MR will result in lower confidence than paper.</td>
<td>Hypothesis not supported; There was no statistically significant main effect of instruction method.</td>
</tr>
<tr>
<td><strong>H8.</strong> Confidence associated with the paper folding procedural task will be influenced by time of recall in that retention-test will result in lower confidence than post-test, followed by immediate recall.</td>
<td>Hypothesis supported; Immediate &gt; Post &gt; Retention</td>
</tr>
<tr>
<td><strong>H9.</strong> Workload associated with the paper folding procedural task will be influenced by instruction method in that AR will result in higher workload MR, followed by paper.</td>
<td>Hypothesis not supported; There was no statistically significant difference between instruction methods.</td>
</tr>
<tr>
<td><strong>H10.</strong> User experience associated with the paper folding procedural task will be influenced by instruction method in that paper will result in a higher pragmatic score than MR, followed by AR.</td>
<td>Hypothesis not supported; There was no statistically significant difference between instruction methods.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>H11.</strong> User experience associated with the paper folding procedural task will be influenced by instruction method in that MR will result in a higher hedonic score than AR, followed by paper.</td>
<td>Hypothesis supported; MR &gt; AR &gt; Paper</td>
</tr>
<tr>
<td><strong>H12.</strong> Trainee reactions associated with the paper folding procedural task will be influenced by instruction method in that paper will result in higher trainee reactions than MR, followed by AR.</td>
<td>Hypothesis partially supported; Significantly higher ratings for paper regarding navigation and ease of use.</td>
</tr>
<tr>
<td><strong>H13.</strong> Relationships between select demographic variables (e.g., spatial visualization abilities; prior XR, video game, and origami experience) and performance will be examined.</td>
<td>Significant, moderate positive correlations between paper, AR, and MR performance and spatial visualization abilities, as well as between paper and AR performance and origami experience.</td>
</tr>
<tr>
<td><strong>H14.</strong> Differences between the dependent variables (e.g., time to completion, perceived task difficulty, perceived confidence) for paper, AR, and MR first and second training sessions will be examined.</td>
<td>Second training sessions were significantly faster and easier for all instruction methods, and participants were more confident in their performance.</td>
</tr>
<tr>
<td><strong>H15.</strong> Differences in strategies for completing the paper folding procedural task between instruction methods will be examined.</td>
<td>Recall strategies did not differ between instruction methods, with the exception of visualizing the virtual cues that were only available during MR training.</td>
</tr>
<tr>
<td><strong>H16.</strong> Differences in likes, dislikes, and recommendations for improvement between paper, AR, and MR will be examined.</td>
<td>Paper was easy to navigate and use, but required participants to take their hands off their work. AR/MR could be used hands-free, but had navigation and hardware issues. AR required more head movement. MR virtual cues were helpful, but found to be obstructive.</td>
</tr>
</tbody>
</table>
Discussion

See Table 13 for a summary of main findings associated with each study measure.

Table 13

Summary of Study Results

<table>
<thead>
<tr>
<th>Measure</th>
<th>Main Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>Decreased performance as time since training increased. No significant difference between instruction methods.</td>
</tr>
<tr>
<td>Time to Completion</td>
<td>Slower completion times during the first training session compared to second. No significant difference between recall sessions or instruction methods.</td>
</tr>
<tr>
<td>Perceived Difficulty</td>
<td>Increased perceived difficulty as time since training increased. No significant difference between instruction methods.</td>
</tr>
<tr>
<td>Perceived Confidence</td>
<td>Decreased perceived confidence as time since training increased. No significant difference between instruction methods.</td>
</tr>
<tr>
<td>Workload</td>
<td>No significant difference between instruction methods.</td>
</tr>
<tr>
<td>User Experience</td>
<td>Overall scores were rated as Excellent for MR, Good for AR, and Below Average for Paper. Instruction methods were comparable in pragmatic quality. MR was highest in hedonic quality, followed by AR, then paper.</td>
</tr>
<tr>
<td>Trainee Reactions</td>
<td>Paper instructions were easier to navigate and use, but were perceived to be less comprehensive and required participants to take their hands off their work. AR/MR instructions were accessible hands-free, but were associated with navigation and hardware issues. AR training required more head movement. MR virtual cues were helpful, but could be obstructive.</td>
</tr>
<tr>
<td>Cybersickness</td>
<td>Concerning oculomotor symptoms, significant disorientation symptoms, and minimal nausea symptoms.</td>
</tr>
<tr>
<td>Preference</td>
<td>AR was the least preferred method of instruction. MR was thought to be more comprehensive and integrative, resulting in reduced perceived cognitive and physical workload. Paper and AR presented training content in a similar manner, but AR involved wearing a cumbersome headset and learning a new method of learning.</td>
</tr>
<tr>
<td>Open-Ended Responses</td>
<td>Common strategies to facilitate procedure recall included utilizing mental images, muscle memory, and the provided diagram of the finished model. Recall strategies did not differ between instruction methods, with the exception of visualizing the virtual cues that were only available during MR training. Most participants believed their recall performance was impacted by how they learned each origami model.</td>
</tr>
</tbody>
</table>
Theoretical Implications

This study contributes to XR training literature by providing insight into objective and subjective differences between paper-, AR-, and MR-based training experiences. This study found that all three training modalities were comparable in regards to performance, completion time, perceived task difficulty, perceived instruction method difficulty, perceived confidence in successfully completing the task, workload, pragmatic quality, and retention recall strategies. Differences were found for user experience and trainee reactions, such that MR received higher user experience ratings and AR was least preferred by participants.

Results of the current study support those presented in the meta-analysis by Kaplan et al. (2021), where the authors concluded that XR training and traditional instruction methods produce an equivalent performance result. The authors also state that if performance between these two types of training is essentially the same, then XR training may be the superior option due to the several advantages of XR training over traditional methods (e.g., reduced costs, safer training, enhanced flexibility).

The current study also supports multiple theories and prior efforts that demonstrate how AR/MR technologies can be leveraged to promote learning. One such theory is the Ganier et al. (2000) Model of Processing Procedural Instructions, which illustrates the amount of attentional switching required to complete procedural tasks. Sources of information in which trainees must manage include the instructions, their mental model of the instructions, their progress, and their prior knowledge, as well as the workspace and equipment state. The need to attend to such a large amount of information can increase trainees’ workload. Because MR can integrate the instructions within the participants’ workspace by anchoring virtual cues onto the physical world, it has the potential to reduce attentional switching and the subsequent workload. The current study supports this notion, and is consistent with findings described by Marner et al. (2013) and Tang et al. (2003). AR workload ratings trended higher
than MR, which can be explained by the greater distance between the AR instructions and the participants’ workspace. As a result of the greater distance, participants reported having to move their head more during the AR training condition in order to reference the instructions. Future work could confirm these subjective reports of increased head movement by using head motion trackers as participants complete training delivered through traditional, AR, and MR means.

Prior work indicates another benefit of AR/MR training is that it can support experiential learning (Asad et al., 2021; Pomerantz, 2019), or “learning by doing” (Gentry, 1990). AR/MR trainees can access supplementary virtual information while interacting with real-world objects (Neumann & Majoros, 1998; Sautter & Daling, 2021), providing trainees with tactile feedback that can be incorporated into their sensorimotor memory storage (Webel et al., 2013). Thus, AR/MR trainees can partake in active, participatory learning opportunities, instead of being limited to observational learning that is common in traditional training (e.g., watching a lecture, video, or demonstration). The idea that AR/MR training can promote sensorimotor memory encoding is supported by the current study. Subjective comments indicated that over half of participants utilized muscle memory as a strategy for recalling the procedure when the instructions were no longer available. However, this strategy was reportedly employed across all three instruction methods - paper, AR, and MR - likely because participants were physically performing the task during all instruction method conditions. Future work could further investigate whether AR/MR training is better at tapping into trainees’ sensorimotor memory storage by comparing AR/MR training to other traditional training methods besides paper-based instructions.

The Method of Loci, a mnemonic device where individuals link information to specific landmarks within a physical space (Maguire et al., 2003; Sternberg & Sternberg, 2012), has also been connected to AR/MR training. Based on this notion, presenting trainees with virtual information that is anchored to the real-world, as in MR, may help trainees memorize information more effectively (Tang et al., 2003).
The results of the current study partially support this idea. Eight participants (27%) commented they visualized the MR virtual cues, even when they were no longer projected to them, as they recalled the procedure from memory. Six of these participants stated this strategy facilitated their ability to recall the procedure, one indicated it did not help their ability to recall the procedure, and one stated it hindered their recall ability. Thus, future research is needed to better understand whether the Method of Loci assists the recall of procedures learned through AR/MR training.

**Practical Implications**

Knowledge of the differences between paper, AR, and MR training can be used to determine which type of training is best suited for a particular use case. Based on the results of the present study, paper instructions may be useful when time available for training is limited, as trainees are more likely to be familiar with this type of instruction. As a result, trainees would not have to expend additional time and resources learning a new way of learning prior to completing their required training. Additionally, the results of this study suggest that it may not be worthwhile to invest in transitioning paper instructions to an AR experience if the AR experience is simply a virtual recreation of the paper experience, unless it is beneficial for trainees to have hands-free access to the training materials so they can keep their hands on task.

Furthermore, participants in the current study subjectively reported that MR may reduce head movements and cognitive workload because it integrated multiple sources of information into one place. Consequently, MR may be suitable for longer training sessions because it may minimize risk of fatigue or injury that could result from having to perform repetitive bodily movements to reference multiple, separate sources of information. However, it would be important to monitor trainees’ cybersickness symptoms as they complete AR/MR training to ensure trainees’ are not experiencing adverse effects. MR virtual content would also have to be designed to reduce obtrusiveness and piloted to confirm it is properly aligned with the physical world. The design of virtual content presented in
AR/MR training experiences should be an iterative process that involves input from an interdisciplinary team that includes instructional designers, human factors practitioners, XR developers, and representative end-users.

If an organization decides to introduce AR/MR training to their workforce, it is recommended that they first allot time for trainees to become more familiar with the AR/MR devices. During an ideal familiarization period, trainees would learn how to adjust the device to promote user comfort (e.g., how to tighten or loosen aspects of the device; how to add or remove accessories) and complete the eye calibration procedure, if one is available, to facilitate accurate presentation of and interaction with virtual content. Trainees should also complete a tutorial to learn how to interact with virtual content. Depending on the input methods accepted by the AR/MR device and application, this tutorial may include learning more about how to use gestures, voice commands, and/or controllers. Trainees should practice utilizing these input methods with tasks similar to what they will be performing during the training session. Examples of such tasks include activating, deactivating, placing, moving, or resizing virtual content. Trainees should also learn how to properly hold and clean the device so it can be sanitized between users. Skipping this familiarization period and simply handing off AR/MR devices to trainees who are inexperienced with using such devices will likely have a negative impact on training effectiveness, user experience, and trainee acceptance of the technology.

Limitations

The current study had several limitations associated with the sample, methodology, and generalizability to other XR devices and training tasks.

**Limitations of Sample.** Participant demographics affiliated with this study may not completely generalize to the general population, as they were primarily college students recruited from Embry-Riddle Aeronautical University. Furthermore, the study’s results may have been impacted by several individual differences, such as learning style, acceptance of XR technologies, motivation to learn and
perform, self-efficacy, and spatial abilities that may not have been captured by the PFT. For example, a few participants described themselves as “perfectionists”, indicating they may have tried harder to complete the task correctly, taken more time to complete the task, or rated themselves lower on the perceived confidence scale.

**Methodology Limitations.** Because this study was conducted in a controlled, laboratory environment, it may have limited ecological validity. This may prevent the results from fully generalizing to real-world training experiences. Additionally, minimizing participant drop outs by reducing the number of study sessions resulted in participants completing their training during a single two-hour session. While participants were offered multiple opportunities to take breaks during the training session, participants could have experienced fatigue during the initial session, negatively impacting their ability to attend to the training materials or recall the procedure from memory. A single, extended training session is also not representative of all training delivery schedules. Moreover, there are limitations regarding how the dependent variables were measured in this study. For one, performance was assessed using a non-validated rubric and individual differences between the three raters who evaluated participants’ completed origami models likely impacted their ratings due to some rubric items yielding more room for subjectivity. Additionally, several dependent variables were collected using self-reported measures. The accuracy of self-reported measures can be hindered by participants forgetting pertinent details or responding in a manner they believe will be viewed favorably by the researchers.

**Limitations in Generalizability to Other Devices and Tasks.** This study only utilized one AR/MR headset. Because different headsets have different attributes (e.g., input methods, hardware features) that can impact user experience and training effectiveness, it may not be appropriate to generalize the findings of the current study to all XR devices used for training. Additionally, the results of this study may be limited to procedural tasks and may not be applicable to cognitive or affective tasks. Also, the task performed in this study was a tabletop task completed by an individual that utilized smaller
equipment, an experience which may differ from tasks that require standing or walking while interacting with larger equipment. This task also utilized a limited range of MR virtual cues (e.g., dotted lines, arrows) that may not be applicable to other tasks.

**Future Research**

Recommendations for future research include replicating the study with a larger, more diverse sample. Individual differences also could be studied to determine what, if any, effect they have on AR/MR training effectiveness and user experience. For example, the current study found significant relationships between spatial visualization ability and task performance, which should be further explored in future research. The use of XR training in a real-world setting and with other training delivery schedules, such as multiple, shorter sessions, could also be investigated. It is recommended for future studies that aim to assess performance to consider how precise accuracy must be measured to determine task success. The need for more precise measures of accuracy may require the use of more objective tools. For example, a more precise measure of accuracy for the procedural task completed in this study would be to evaluate completed origami models using rulers and protractors. The current study could also be replicated using other types of tasks (e.g., cognitive, affective), MR virtual cues, and XR devices used for training, including other AR/MR headsets, AR/MR handheld devices, and VR devices. Other procedural tasks could be also examined, especially those that involve standing or walking to interact with larger or smaller objects, or those that require collaboration with others.
References


https://doi.org/10.1002/9781119636113.ch16


Origami.me. (n.d.). *Diagrams*. https://origami.me/diagrams


103


Appendices

Appendix A

*Origami Models and Sources*

Practice Models

- **Cup**
  - Adapted from: https://origami.me/cup/

- **Sailboat**
  - Adapted from: https://origamiusa.org/files/sailboat.pdf

- **Dog**
  - Adapted from: https://origamiusa.org/files/traditional_dog.pdf

Experimental Models

- **Samurai Helmet**
  - Adapted from: https://origamiusa.org/files/diagrams_pdf/samurai_helmet.pdf

- **Cicada**
  - Adapted from: https://origamiusa.org/files/diagrams_pdf/cicada.pdf

- **Necktie**
  - Adapted from: https://origami.me/necktie/
Appendix B

Demographic Questionnaire

1. Participant ID: ______
2. Age: ______
3. Gender:
   - Male
   - Female
   - Other ______
4. Highest level of completed education:
   - Some high school
   - High school diploma or equivalent
   - Some college
   - Undergraduate degree
   - Graduate degree
5. Occupation (if you are currently a Student, please put Student and write your major): _____
6. Do you own any augmented, mixed, or virtual reality devices? For example, Oculus Rift, HTC Vive, Microsoft HoloLens, AR-enabled smartphone.
   - Yes [Advance to Question 7]
   - No [Advance to Question 8]
7. What augmented, mixed, or virtual reality devices do you own? Select all that apply.
   - Oculus Rift
   - Oculus Quest
   - HTC Vive
   - Google Cardboard
   - Google Glass
   - Microsoft HoloLens 1
   - Microsoft HoloLens 2
   - Other _____
8. Have you used any augmented, mixed, or virtual reality devices before this study? For example, Oculus Rift, HTC Vive, Microsoft HoloLens, AR-enabled smartphone.
   - Yes [Advance to Question 9]
   - No [Advance to Question 12]
9. What augmented, mixed, or virtual reality devices have you used before this study? Select all that apply.
   - Oculus Rift
   - Oculus Quest
   - HTC Vive
10. Approximately how many hours have you used augmented/mixed reality (AR/MR) headsets (e.g., Microsoft HoloLens, Magic Leap, Google Glass) before this study?
   - Less than 1 hour
   - 1 to 4 hours
   - 5 to 9 hours
   - 10 to 19 hours
   - 20 to 29 hours
   - 30 to 39 hours
   - More than 40 hours

11. Approximately how many hours have you used virtual reality (VR) headsets (e.g., Oculus Quest, HTC Vive) before this study?
   - Less than 1 hour
   - 1 to 4 hours
   - 5 to 9 hours
   - 10 to 19 hours
   - 20 to 29 hours
   - 30 to 39 hours
   - More than 40 hours

12. Do you play video games?
   - Yes [Advance to Question 13]
   - No [Advance to Question 17]

13. What type of video game player do you consider yourself to be?
   - Newbie/Novice
   - Casual
   - Mid-core
   - Hardcore/Expert

14. On average, approximately how many hours do you spend playing video games per week?
   - Less than 1 hour
   - 1 to 4 hours
   - 5 to 9 hours
   - 10 to 19 hours
   - 20 to 29 hours
   - 30 to 39 hours
15. Which of the following devices do you frequently use to play video games? Select all that apply.
   - A computer device (e.g., laptop, desktop)
   - A console device (e.g., Xbox One, PlayStation 4, Nintendo Switch)
   - A handheld gaming device (e.g., Nintendo DS, PlayStation Vita)
   - A mobile device (e.g., smartphone, tablet)
   - A headset (e.g., Oculus Quest, HTC Vive)

16. Which of the following video game genres do you frequently play? Select all that apply.
   - Action (e.g., Halo, Call of Duty)
   - Adventure (e.g., Resident Evil, Grand Theft Auto)
   - Driving (e.g., Forza, Mario Kart)
   - Educational/Edutainment (e.g., Math Blaster, Professor Layton)
   - Fighting (e.g., Soul Caliber, Mortal Kombat)
   - Fitness (e.g., Wii Fit, Your Shape: Fitness Evolved)
   - Music/Dance (e.g., Guitar Hero, Just Dance)
   - Puzzle/Card (e.g., Tetris, Solitaire)
   - Retro/Classic (e.g., Pacman, The Original Donkey Kong)
   - Role Playing (e.g., Elder Scrolls, World of Warcraft)
   - Simulation (e.g., The Sims, Spore)
   - Social/Social Network (e.g., Farmville, Candy Crush)
   - Sports (e.g., Madden NFL, FIFA)
   - Strategy (e.g., Civilization, Starcraft)

17. How familiar are you with origami (i.e., following diagrams/instructions to fold paper into art)?
   - Extremely familiar
   - Moderately familiar
   - Somewhat familiar
   - Slightly familiar
   - Not at all familiar

18. How proficient are you in completing origami models?
   - Extremely proficient
   - Moderately proficient
   - Somewhat proficient
   - Slightly proficient
   - Not at all proficient
Appendix C

**Paper Folding Test**

**Paper Folding Test—Vz-2-BRACE**

In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line. The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point. One of the five figures on the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and draw an X through that figure.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper).

![Sample Problem Diagram]

The correct answer to the sample problem above is C and so it should have been marked with an X. The figures below show how the paper was folded and why C is the correct answer.

![Folding & Unfolding Diagrams]

In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

Some of the problems on this sheet are more difficult than others. If you are unable to do one of the problems, simply skip over it and go on to the next one.

You will have three minutes for each of the two parts of this test. Each part has one page. When you have finished Part One, STOP. Please do not go on to Part Two until you are asked to do so.

**DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO**

112
**PART ONE (3 MINUTES)**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

113
PART TWO (3 MINUTES)

11

12

13

14

15

16

17

18

19

20

A

B

C

D

E
### Appendix D

**Performance Scoring Rubrics**

<table>
<thead>
<tr>
<th><strong>Cicada</strong></th>
<th><strong>Yes</strong></th>
<th><strong>No</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Does it approximate the example model?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[If “Yes”, complete 2-6; If “No”, finished]</td>
<td>No</td>
</tr>
<tr>
<td>2 Is it symmetrical?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>3 Are the three front body angles proportionally spaced?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>4 Are the three points of the front body angles aligned?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>5 Are both of the wings folded such that they create parallel lines?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>6 Are the tops of the wings covered by the front body pieces?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>7 On the back side, are the wings separated by a small gap?</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Samurai Helmet</strong></th>
<th><strong>Yes</strong></th>
<th><strong>No</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Does it approximate the example model?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[If “Yes”, complete 2-6; If “No”, finished]</td>
<td>No</td>
</tr>
<tr>
<td>2 Is it symmetrical?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>3 Does it operate like a hat (bottom opens as if it could be placed on a head)?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>4 Is the front band present?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>5 Are both horns present and folded to the front?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>6 Are the horn points aligned horizontally?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>7 Is the front triangle centered with its point below the horns?</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Necktie</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>1 Does it approximate the example model?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[If “Yes”, complete 2-6; If “No”, finished]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Is it symmetrical and continuous?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Is the knot present and does it make a pocket?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Is the knot six-sided?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 On the back side, are the edges folded to the center?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 On the back side, is the triangle at the top present with its point</td>
<td></td>
<td></td>
</tr>
<tr>
<td>covered by other folds?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 On the back side, are the top triangle pockets present?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix E

Single Ease Question

Seven-point rating scale response (1 = Very Difficult; 7 = Very Easy)

[Question asked following completion of each origami model:] Overall, how difficult or easy was it to put together the model?

[Question asked following completion of each instruction method:] Overall, how difficult or easy did you find using this method of instruction?
Appendix F

Confidence

Seven-point rating scale response (1 = Not at all Confident; 7 = Very Confident)

Overall, how confident are you that you successfully completed this origami model?
Appendix G

Raw NASA-TLX

Mental Demand
How mentally demanding was the task?

Very Low

Very High

Physical Demand
How physically demanding was the task?

Very Low

Very High

Temporal Demand
How hurried or rushed was the pace of the task?

Very Low

Very High

Performance
How successful were you in accomplishing what you were asked to do?

Perfect

Failure

Effort
How hard did you have to work to accomplish your level of performance?

Very Low

Very High

Frustration
How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low

Very High
### Appendix H

**User Experience Questionnaire - Short Version**

<table>
<thead>
<tr>
<th>term</th>
<th>scale</th>
<th>adjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>obstructive</td>
<td>O O O O O O O O</td>
<td>supportive</td>
</tr>
<tr>
<td>complicated</td>
<td>O O O O O O O O</td>
<td>easy</td>
</tr>
<tr>
<td>inefficient</td>
<td>O O O O O O O O</td>
<td>efficient</td>
</tr>
<tr>
<td>confusing</td>
<td>O O O O O O O O</td>
<td>clear</td>
</tr>
<tr>
<td>boring</td>
<td>O O O O O O O O</td>
<td>exciting</td>
</tr>
<tr>
<td>not interesting</td>
<td>O O O O O O O O</td>
<td>interesting</td>
</tr>
<tr>
<td>conventional</td>
<td>O O O O O O O O</td>
<td>inventive</td>
</tr>
<tr>
<td>usual</td>
<td>O O O O O O O O</td>
<td>leading edge</td>
</tr>
</tbody>
</table>

*Note.* The first four items correspond to Pragmatic Quality, while the last four items correspond to Hedonic Quality.
Appendix I

Trainee Reactions

Five-point Likert scale responses (1 = Strongly Disagree; 5 = Strongly Agree)

1. The training content was clear.
2. I could easily understand the training content.
3. I was able to navigate through the training content.
4. I found the training content easy to use.
5. I was satisfied with the presentation of the training content.
6. I had a positive learning experience.

Open-ended responses

7. What did you like best about this training?
8. What did you dislike about this training?
9. How would you improve this training?
## Appendix J

### Outliers

The following data points were identified as greater than ±3 standard deviations from the mean.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Condition</th>
<th>Data Point Identified as Outlier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completion Time</td>
<td>Paper, Retention</td>
<td>1 completion time of 481 seconds</td>
</tr>
<tr>
<td></td>
<td>MR, Training 1</td>
<td>1 completion time of 437 seconds</td>
</tr>
<tr>
<td></td>
<td>MR, Post</td>
<td>1 completion time of 430 seconds</td>
</tr>
<tr>
<td>Perceived Difficulty</td>
<td>Paper, Immediate</td>
<td>1 response of “1”</td>
</tr>
<tr>
<td></td>
<td>Paper, Training 2</td>
<td>1 response of “2”</td>
</tr>
<tr>
<td></td>
<td>AR, Training 2</td>
<td>1 response of “1”</td>
</tr>
<tr>
<td></td>
<td>AR, Immediate</td>
<td>1 response of “1”</td>
</tr>
<tr>
<td></td>
<td>MR, Training 2</td>
<td>1 response of “2”</td>
</tr>
<tr>
<td>Workload</td>
<td>Paper</td>
<td>1 response to Performance as “21”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 response to Frustration as “20”</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>1 response to Performance as “19”</td>
</tr>
<tr>
<td></td>
<td>MR</td>
<td>1 response to Temporal as “14”</td>
</tr>
<tr>
<td>User Experience</td>
<td>Paper</td>
<td>1 response to Item 3 as “1”</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>1 response to Item 3 as “1”</td>
</tr>
<tr>
<td></td>
<td>MR</td>
<td>1 response to Item 6 as “4”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 response to Item 8 as “3”</td>
</tr>
<tr>
<td>Trainee Reactions</td>
<td>Paper</td>
<td>1 response to Item 3 as “4”</td>
</tr>
<tr>
<td></td>
<td>AR</td>
<td>1 response to Item 3 as “2”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 response to Item 4 as “1”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 response to Item 6 as “2”</td>
</tr>
<tr>
<td></td>
<td>MR</td>
<td>1 response to Item 1 as “2”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 response to Item 2 as “2”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 responses to Item 6 as “2”</td>
</tr>
</tbody>
</table>