Improving Usability for Mobile eXtended Reality (XR): A Case Study using Tactical Combat Casualty Care Training and Readiness to Measure Training Effectiveness

Betsy Guzmán Laxton, JoAnn Archer
Design Interactive, Inc.
Orlando, FL
Betsy.Laxton@designinteractive.net, JoAnn.Archer@designinteractive.net

ABSTRACT

Military leaders have identified zero preventable deaths as a high priority for improving the survival of those injured during combat. The Department of Defense (DoD) has focused on initiatives to improve Tactical Combat Casualty Care (TCCC) training such as the Deployed Medicine platform. Limitations with existing training include the inhibitive costs of realistic, interactive medical manikins and eXtended reality (XR) headsets, and the lack of opportunities to practice and perform trauma care. Due to high throughput required and time necessary to complete this training, service members find it difficult to attend and maintain competency in the TCCC skill set. Accessible, goal-specific, and guided training on mobile devices with XR simulations has the potential to improve survivability statistics for preventable deaths.

Training technologies such as virtual reality (VR), augmented reality (AR), mixed reality (MR) or the combination of all three under XR is emerging as a realistic, low-cost immersive solution with potential in improving and supporting performance of skills. Moreover, mobile XR, which utilizes tablet and cell phone devices, presents high fidelity augmentations to real-world elements offering trainees realistic environments for prescribed learning opportunities in an accessible and familiar manner. By utilizing activities and assessments developed especially for mobile XR to assess skills and proficiency, trainees can interact with and embody the real and virtual environment through hands on practice that facilitates cognitive, psychomotor, and affective learning. This paper presents the results from a case study analyzing usability data collected at multiple military training centers of lessons developed in a mobile XR TCCC application. Those data were used to iteratively create improved XR interactions and usability of the application in preparation for a Training Effectiveness Evaluation (TEE) to. minimize any usability-related effects on the performance measures.

ABOUT THE AUTHORS

Besty Guzmán Laxton is a Research Associate at Design Interactive, Inc. with an industrial engineering and sociology background. She has 4 years of experience conducting Human Factors research with a focus on eXtended Reality (XR) and over 10 years of project management experience in the medical practice field focused on creating efficient work systems. Her combined expertise contributes to the development and usability of XR training applications for an improved user experience.

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

Improving Usability for Mobile eXtended Reality (XR): A Case Study using Tactical Combat Casualty Care Training and Readiness to Measure Training Effectiveness

Betsy Guzmán Laxton, JoAnn Archer
Design Interactive, Inc.
Orlando, FL
Betsy.Laxton@designinteractive.net, JoAnn.Archer@designinteractive.net

INTRODUCTION

The Department of Defense (DoD) has prioritized initiatives that improve battlefield survival outcomes. Platforms for Tactical Combat Casualty Care (TCCC) training such as Deployed Medicine, the medical simulation training environment (mSTE) and the Squad Virtual Trainer and Integrated Visual Augmentation System (SSVT IVAS), are supported as venues for improving medical training with the end goal of minimizing preventable deaths on the battlefield. Current training systems often come with the inhibitive costs of realistic, interactive medical manikins and live training exercises that lack the opportunity to practice and perform trauma care independently to obtain, retain, and sustain TCCC skills. Training technologies such as virtual reality (VR), augmented reality (AR), mixed reality (MR) or the combination of all three under eXtended Reality (XR) are emerging as realistic, low-cost immersive solutions with potential in improving and supporting performance of skills. XR technology presents high fidelity augmentations to real-world elements, offering trainees realistic environments and learning opportunities. By utilizing activities and assessments developed especially for XR to assess skills and proficiency, trainees can interact with and embody the real and virtual environment through hands on practice that facilitates cognitive, psychomotor, and affective learning. When utilizing XR on mobile devices (mobile XR) such as IOS or Android tablets and cell phones, the opportunity to access TCCC training during formal course enrollment, as refresher training, or in preparation for evaluation is substantially increased resulting in improved forward operating readiness for the war fighter.

This paper presents an analysis of data collected from a case study to evaluate the usability of a mobile XR application for TCCC training. The case study consists of four usability evaluations conducted at multiple military training centers and two expert heuristic evaluations of TCCC lessons. Results from the case study were used to improve XR interactions and usability of the application in preparation for a Training Effectiveness Evaluation (TEE) of the mobile XR application. The TEE will test improved performance for trainees who undertake classroom training with the XR mobile application as compared with trainees who undertake the classroom training using traditional materials. The goal is that usability related effects will be minimized on the performance measures of the TEE.

BACKGROUND

Current TCCC Training

Typical battlefield care providers include Combat Medics and Combat Life Savers (CLS). Combat Medics are military personnel that have completed at minimum 16-weeks of training to provide first aid and frontline trauma care (Moralez et al., 2021). They receive advanced training in the form of simulations to provide the necessary realism of battlefield stress, complexity, and dynamics (Fitzhugh, 2015). Continuous training and advancement of their skills is the norm, making them the definitive lifesaving resource in combat. On the other hand, CLS are intended to provide an intermediate step between the basic life support skills taught to every Soldier and Combat Medics. CLS typically receive only 40 hours of instruction prior to certification. Thus, they can often be ill prepared to provide appropriate and timely battlefield trauma care for Soldiers with survivable injuries due to limited opportunities to practice under the same contextually representative conditions as Combat Medics. The extent of their tactical trauma care training is inhibited by the prohibitive cost of live training, medical manikins, and a lack of anthropically diverse manikins or casualty simulators that can present with physiological realism of the types of injuries experienced on the battlefield. Combat lifesaving is a complex, high tempo process situated in and demonstrably interwoven within the battlefield context. The battlefield is highly fluid, filled with dynamic noises (e.g., explosions, weapon fire, vehicle engine, etc.), taxing environmental factors (e.g., heat, cold, humidity, etc.), visual confusion (e.g., due to fog, smoke, sandstorms, etc.), and other such factors that present stressors that directly impact operational performance (Merlo, 2018). These contextual factors may hinder a CLS's ability to rapidly scan a scene, focus on relevant cues, develop a casualty response plan, and administer care. It is critically important to prepare CLS to perform potentially lifesaving care in a variety of challenging battlefield environments. Providing such training without considering the impact of contextual factors would be expected to impede training transfer. This training gap limits access to repeatable and representative training examples when military personnel are not deployed. Accessible, goal-specific, and guided just-in-time training for trauma resuscitation has the potential to improve survivability statistics in complex and time-sensitive situations experienced by non-medic (bystander) military personnel and civilians. These first responders have few opportunities to practice under conditions representative of battlefield stress, complexity, and dynamics. XR technology can provide CLS with just in time training that is contextually relevant through scenarios that require trainees to adapt to the demands of the battlefield while using cognitive, psychomotor, and affective skills.

Case Study Application

The mobile XR application used for the case study is AUGMED® (Augmented Reality for Adaptive, Effective Learning and Execution of Medical Care) Mobile, a medical XR tool that is currently in development by the DoD to address TCCC training gaps. The case study application is an adaptive, modular, immersive, and extensible XR training system geared toward CLS providers who are learning the TCCC curriculum. The key aim of this system is to implement contextualization and embodied learning within an XR environment such that trainees reach proficiency faster than standard training methods across all learning skills (cognitive, psychomotor, and affective) thereby increasing trainees' understanding of material and decreasing the overall training time required to reach a competent level. The system leverages XR technology to bridge the gap from traditional classroom-based learning to hands-on skills training. It is designed to effectively deliver spatialized instructional content and contextualized, realistic training cues to learners. However, a crucial first step to developing XR training solutions that leverage principles of contextualization and embodied cognition to enhance training is to evaluate usability of novel contextualized design elements and embodied interactions afforded by XR. If the embodied interaction techniques featured in the system are too cumbersome for trainees to intuitively use, then transfer of training is likely to be diminished. Careful examination of these features can highlight positive and negative experiences in XR, possible improvements to overall usability, and future directions for evaluating the application of contextualization and embodied cognition principles into the design of XR training solutions for Combat Lifesavers.

Currently, the mobile XR application used for the case study consists of 14 lessons covering the MARCH (Massive Hemorrhage, Airway, Respiration, Circulation, Head Injury/Hypothermia) continuum. Lessons were developed based on adaptation strategies designed and developed for XR consumption (Stanney et al., 2021) and are organized by the following lesson types: Symptom and Injury, Tools and Treatment, Skills Practice, Scenario, and Assessment Practice. A total of seven activities and assessments were developed for pedagogically driven contextualization and embodied cognition. Each activity has been carefully crafted to afford a variety of contextually rich and embodied formative assessments within a room-scale 3D training experience when used in the mobile XR form factor. The activities and assessment include the following:

1. **Discovery**: The trainee is presented with a virtual casualty that features various points of interest that are highlighted through spatial iconography. Trainees are prompted to physically move close to areas of interest to learn more information. As a trainee approaches an area of interest, contextually relevant visual and auditory information is revealed.
2. **Locate and Select**: Spatialized iconography is appended to the casualty and the trainee is prompted to select one or more areas of interest to answer a question about the situation. This requires the trainee to physically look at a specified area while focusing visually before deciding on the correct location.
3. **Matching**: The trainee is presented with a 3D shelf of various objects. Objects may take the form of 3D models, images, or labels. The trainee is prompted to drag-and-drop an object from the shelf and physically match it to its associated location in space per the training content.
4. **Question and Answer**: Audiovisual questions are presented to the trainee for a quick assessment. The questions prompt the trainee to respond based on what they are perceiving in 3D space.
5. **Action Plan** The action plan includes a list of items that need to be sorted into a correct sequence. The action plan can be presented to a trainee in an incomplete or partially complete state depending on proficiency level. After the action plan is completed, the trainee can carry out the plan of action.

6. **Timed Activity**: Time to task completion is captured by the trainee selecting start to begin a countdown which reflects the time within which they need to complete an intervention. This is a simple interaction design that introduces contextually relevant visual and auditory time stress while capturing trainee performance.
7. **Navigate Activity:** The navigate activity occurs within the 3D environment, where the trainee must assess the options for where to move based on the scenario. The trainee makes their decision based on their understanding of the situation and then physically moves to the corresponding location.

The application used for the case study was deployed to several device types including iPhone 14, iPhone 14 Pro, iPad Air, iPad, Pixel 6A, and Pixel 6.

Technical Effectiveness Evaluation

TEE's often rely on final performance outcome measures to address the value of a training intervention. The focus on one measure of effectiveness does not consider the different levels of performance including acquiring targeted knowledge, skill acquisition, or change in behavior (Cohn et. al, 2007). A TEE is usually conducted after the training system has been completed with the focus on performance outcomes. Performance measures that focus on final outcomes without analyzing cognitive, psychomotor, and affective learning miss the opportunity to include an analysis of how the training system is contributing to learning. By incorporating a life-cycle approach to evaluating the effectiveness of a training system, the different metrics can be analyzed to provide a holistic view of the training system's contribution to learning and outcomes.

Using a life-cycle approach, a training system should be developed with both the theory and practice integrated into the function of the system. The approach entails starting with a training needs analysis (TNA) and Sensory Task Analysis (STA) which focus on identifying the appropriate goals and objectives that the system is addressing (Cohn et al., 2007). The TNA and STA are analyzed to identify the appropriate cues and how they are acquired. Those results are used to create a mix of methods that will target cognitive, psychomotor, and affective learning. The system design is based on the research to support learning as well as meeting operational functions and coordination requirements. By understanding the needs that the system must address, the performance outcomes would be constructed to address those metrics. A thorough TNA and STA that inferred the system design for the XR application used for this case study was completed prior to the development of the system.

Cohn et al. (2007) offer a framework that ensures the question of whether or not the training system affords effective learning. The basic tenets of the framework cover a Theoretical TEE, a Trainee Performance Evaluation, a Training Transfer Evaluation, and a Transfer Efficacy analysis. By using this framework and incorporating Kirkpatrick's model for evaluation training effectiveness (Kirpatrick, 1959), the system analyzes the system for learning from multiple angles. At the beginning of the design process, theoretical TEE looks at how the system incorporates the cues that are part of the learning goal. Usability testing can be utilized to inform the theoretical TEE and contributes to the TEE framework. The results provide data in line with Kirkpatrick's level 1 to indicate the training system's relevance to the user. Training performance evaluations help to address Kirkpatrick's evaluation level 2 which address metrics. Using Bloom's taxonomy, learning can be analyzed along 3 dimensions (cognitive, psychomotor, and behavior) (Bloom et al., 1956). The training transfer evaluation would align with Kirkpatrick's level 3, behavior, to determine how well the training technology tracks with operational readiness and mission performance. The transfer efficacy would be achieved through results, Kirkpatrick's level 4, which would indicate the impact of the training system on the mission. The iterative approach to user testing as described in the case study can help to identify gaps in requirements that were identified as part of the mobile XR application task analysis and reported as part of the training system requirements.

METHODOLOGY

Participants and Materials

Data for the case study were collected from four user evaluations with end users at military medical training facilities and during two rounds of expert heuristic evaluations. For user testing, participants were recruited to voluntarily participate in the usability evaluation of the mobile XR application. All had normal or corrected to normal vision and were 18+ years old. Participants were recruited from a CLS class, ASM class, from the instructor pool or other administrative personnel who were available during the testing period. Depending on the time available with the participants, 1-5 lessons were reviewed with the group. The participants were equipped with either an iPhone, an iPad, or an Android cellular phone to test a specified CLS lesson using the mobile XR application selected for the case study. Expert heuristic evaluations were completed using a modified version of "The Derby Dozen: an AR/MR Usability Heuristic Checklist," (Derby Dozen) (Derby, 2023) for two lessons at a time. This heuristic checklist was chosen because it has been empirically validated and identifies a comprehensive list of usability issues that could occur with a variety of AR/MR applications and devices (including both HWD and mobile devices).

Participants at the four military medical training facilities were asked to complete the System Usability Score (SUS) survey, respond to open-ended questions, and rate the system features after finishing a mobile XR application lesson. The SUS is an industry standard for research on usability, there are 10 questions, with 5 response options ranging from Strongly Disagree to Strongly Agree to provide an overview of the system's usability (Sauro, 2011). Right skewness of the positively coded questions and left skewness of the inverse coded questions indicate a favorable view of the system and an overall score of 68 or higher is considered average to above average usability (Brook, 1986). The open-ended questions were developed for the purpose of getting more details regarding the participants' perception of the mobile XR application. Similarly, the ability to score the system features separated out content issues from the system functions. The participants were presented with statements regarding the different system features and a 5 response options ranging from Strongly Disagree (1) to Strongly Agree (5). Additionally, the research team collected observational data regarding user interaction.

For each of the two rounds of heuristic evaluations, there were two expert evaluators who each evaluated two lessons at a time. There was one lesson that was evaluated by both experts for a total of 7 lessons that were evaluated. The original Derby Dozen checklist includes 12 heuristics which were adapted to omit heuristics that were not applicable, for a final checklist consisting of 8 heuristics. The modified checklist was used to evaluate the mobile XR application.

Procedures

Testing was conducted in a room separate from offices and classrooms in all cases. Participants came into the testing room and were greeted by the researchers. The participant was given information about the purpose of the mobile XR application and what type of lesson they were going to complete. While undertaking the lesson, participants were asked to think aloud as they progressed. The researcher made note of any comments and feedback and recorded observations related to how the user interacted with the system. Participants were also informed that there would be a short survey after their experience with the mobile XR application.

Results reported below only include data reported for the question, omitting from the calculations any responses that were left blank.

RESULTS

Usability Evaluation One

The first user evaluation was conducted at MacDill Air Force Base 6th Medical Group with a CLS class. Researchers conducted usability testing on two (2) crawl level lessons, Massive Hemorrhage Scenario 1, a scenario-based lesson, and Massive Hemorrhage Symptoms & Injuries, a didactic, non-scenario lesson, during a CLS class. Researchers observed users' actions, noted their system interactions, and recorded user comments for analysis. There was a total of 14 participants who completed 27 responses (one participant only completed one lesson). The SUS data was not completed for 2 lessons and were omitted from the results in Table 1. The overall SUS included all users for both lessons over the two days and had an average of 70.5 (SD=13.13). This showed that in general, the users felt that the mobile XR application had above average usability.

Table 1 SUS results for first user evaluation

	First User Evaluation		
Lesson	**SUS**	**SD**	**N**
Massive Hemorrhage Scenario 1	72.7	15.89	13
Massive Hemorrhage Symptoms & Injuries	68.1	9.42	12
Instructors	76.1	11.40	9

Trainees	67.3	13.31	16
Overall	***70.5***	***13.13***	***25***

Users included both instructors of the TCCC CLS class as well as the trainees. The SUS scores were analyzed by user type to further explore the usability. The instructor scores showed much higher scores for the positively coded questions which translated into a higher average SUS of 76.1 (SD=11.40) than the trainee SUS of 67.3 (SD=13.13). This indicates that the usability for the end-user of the product still needs to be improved. The instructors were familiar with the content reviewed in the lessons, but trainees were not as familiar. The trainees were learning the material and how to operate the application at the same time which may have affected the overall usability.

When comparing the SUS scores between the two crawl level lessons the first lesson, Massive Hemorrhage Scenario 1, was the users' first introduction to the mobile XR application. The average SUS for this lesson was 72.7 (SD=15.89). The SUS for the second lesson, Massive Hemorrhage Symptoms & Injuries, was 68.1 (SD=9.42) notably due to the increase in length, complexity, and hands-on training.

In general, the mobile XR application was in the average to above average category for usability. The data showed that trainees may need more direction on how to proceed since the technology and the content are relatively new to them based on the responses to the inversely coded questions. The SUS results also indicate that shorter, focused lessons resonate more with the users.

Users were also asked to rate statements regarding the system's features. Most features were well understood, as indicated by average scores of 4 and higher and an average of 4.0 (see Table 2). The two statements that scored 3.78 and 3.74 indicate that the users were unsure of how to interact with the system and the 3D content. This finding supports the general trends from the SUS where the instructors found the system more usable since they have more experience with the content and were only learning the system for the first time and that the first lesson, which was shorter and simpler, was considered more usable. The one statement with a score of 2.15, "I had trouble placing the virtual 3D model," was inversely coded and thus the lower average for this statement is considered more favorable.

Table 2 Average scores for system features for all evaluations

	Evaluation			
System Feature Statement	**First**	**Second**	**Third**	**Fourth**
The goal of the lesson was clear	4.3	4.5	4.4	4.3
I had trouble placing the virtual 3D model*	2.1	2.0	2.1	2.9
Instructions are easy to read	4.5	4.3	4.6	3.7
Instructions are easy to hear	4.6	4.5	4.7	3.8
Instructions are easy to follow	4.1	4.4	4.2	3.6
Activities (Discovery, Matching, Locate & Select, Q & A) are easy to use	4.0	4.3	4.3	3.4
Activities are easy to understand	4.1	4.5	4.4	3.8
It was easy to navigate through the lessons	4.1	4.3	4.3	4.0
I easily understood how to interact with the system	3.8	3.5	4.1	3.7
Interacting with 3D content felt intuitive	3.7	3.7	4.4	3.8
System feedback was immediate and helpful	4.0	4.3	4.5	3.9
Medically relevant details are easy to see	4.2	4.6	4.7	3.8
Average	***4.0***	***4.1***	***4.2***	***3.7***

**The second statement in system features questionnaire was inversely coded and had a low score throughout testing indicating that user generally understood how to place the 3D model.*

The analysis from the observation data showed that there was some hesitancy with how to proceed with the matching and locate and select activities. These activities encourage movement by embodying the learned tasks to identify and apply the knowledge. This high level of interaction is significantly different from the expected functionality of typical mobile applications. Another observation noted is that users responded positively to the scenario in the first lesson and were interested in having more exposure to the application due to the high level of engagement.

User Evaluation Two

The second user evaluation was conducted during a one-day ASM class with the 6th Medical Group at MacDill AFB. Researchers conducted usability testing on one crawl scenario level lesson, Massive Hemorrhage Scenario 1, which had been previously tested with CLS users. Users included both instructors and trainees.

Table 3 SUS results for second user evaluation

	Second User Evaluation		
Lesson	**SUS**	**SD**	**N**
Massive Hemorrhage Scenario 1	82.7	9.78	11
Instructors	73.3	8.04	3
Trainees	86.3	8.13	8
Overall	***82.7***	***9.78***	***11***

The overall SUS for the second user evaluation, shown in Table 3, included all users and had an average of 82.7 (SD=9.8). This showed that in general, the users felt that the mobile XR application had above average usability. The SUS scores were separated by type of user to further explore the usability across the different user types. The average SUS for instructors was 73.3 (SD= 8.04) and the average trainee SUS was 86.3 (SD=8.13). This indicates that the usability for the end-user ranged from good to excellent.

Most system features were well understood, as indicated by average scores of 4 and higher and an average of 4.1 (see Table 2). The two statements that scored 3.55 and 3.73 indicate that the users were unsure of how to interact with the system and the 3D content.

The analysis of observational data showed that there was some uncertainty with how to interact with the matching and the locate and select activities, specifically users wanted to slide the tiles into place rather than tap to select response locations. User comments indicated that they were intrigued and would be interested in interacting with additional lessons.

Overall, results from the second user evaluation during the ASM course were consistent with the first. The researchers observed that the embodiment was more pronounced for the actions associated with the discovery and the locate and select activities. These activities encourage movement centered around the casualty and encourage the user to scan the surrounding virtual environment for greater understanding of the situation. These findings underscore the interest in activities that engage the user with the casualty and environment.

User Evaluation Three

The third user evaluation occurred at the Defense Medical Readiness Training Institute at Joint Base San Antonio with TCCC instructors and entailed usability testing two new lessons, the crawl scenario level lesson Airway Tools & Treatment and the run level lesson, Massive Hemorrhage Scenario 3. Researchers also conducted usability testing on lessons that were updated based on previous user testing feedback, including the Massive Hemorrhage Scenario 1 and Massive Hemorrhage Symptoms & Injuries. Due to the feedback from the first user evaluation to keep crawl level lessons focused on fewer lesson objectives, the hemostatic dressing content from the Massive Hemorrhage Symptoms and Injuries lesson was moved to its own lesson, Hemostatic Dressing Tools & Treatment, a crawl, non-scenario lesson which was also user tested as part of user evaluation three.

The participants for the user testing were largely made up of instructors familiar with the TCCC Curriculum and two administrators. There were no trainees among the participants, so lessons were not analyzed by user type. There were 15 unique participants, two of which were not instructors. Nine of the participants tested two different lessons. There was a total of 26 responses collected for the five lessons. The new Massive Hemorrhage Scenario 3 lesson was emphasized since this was the first run lesson developed. The SUS scores for all the lessons are listed in Table 4.

Table 4 SUS results for third user evaluation

	Third User Evaluation		
Lesson	**SUS**	**SD**	**N**
Massive Hemorrhage Scenario 1	83.3	14.43	3
Massive Hemorrhage Symptoms & Injuries	84.2	14.22	3
Hemostatic Dressing Tools & Treatment	72.5	2.5	3
Airway Tools & Treatment	85.8	6.29	3
Massive Hemorrhage Scenario 3	81.1	12.44	9
Overall	***81.3***	***11.14***	***21***

The overall SUS, shown in Table 4, included 21 reported responses for all the lessons and had an average SUS score of 81.3 (STD = 11.14). This indicated that users think it is an excellent system that is effective, efficient, and easy to use**.** The SUS for the Massive Hemorrhage Scenario 3 included 9 responses for this specific lesson and had an average of 81.1 (STD=12.4). This indicates that the users thought this lesson was excellent. The other lessons had fewer than 5 responses each and Table 4 shows the SUS scores for each lesson.

In general, the mobile XR application was considered excellent by the users. The open-ended questions showed that the users liked the interactions in the application, the detailed animations, and the ability to go at their own speed. The interactions they liked least included having to provide several inputs before moving to the next step, and not understanding the mechanics of how to navigate certain aspects of the activities in the application. Specifically, the Hemostatic Dressing Tools & Treatment was the lowest scoring for SUS. Responses to the open-ended questions indicated users noted that a bigger screen would be helpful in seeing all the options available and there seemed to be some uncertainty about how to interact with the casualty. Although the mobile XR application was designed for the ease of mobility offered by a phone, the lessons can also be viewed on a tablet which would provide a larger field of view. The general feedback regarding interacting with the application, such as minimizing the number of buttons that will advance to the next step or activity and adding more visual prompts to help the user advance, will be addressed through feature enhancements to make it easier for all users to use the application. Most features were well understood, as indicated by scores of 4.0 or higher and an average of 4.2 (see Table 2).

Analysis of the observation data showed that the users tried to select a location for the navigation activity rather than moving to the location. This indicates that the users need to have a better understanding of the XR interactions such that they know to move to advance through an activity rather than passively selecting the response.

User Evaluation Four

The fourth user evaluation was conducted at MacDill AFB 6th Medical Group with a CLS class and included usability testing for four crawl, non-scenario lessons: Respiration Symptoms and Injuries lesson (new), Respiration Chest Seal Tool and Treatment lesson (new), Respiration Chest Seal Skill Practice lesson (new), and Airway Tools & Treatment lesson (previously tested at the third user evaluation). Researchers also conducted usability testing on one new run scenario lesson, Respiration Assessment Practice 2 lesson.

The participants for the user testing were largely made up of students attending a CLS class. Each participant tested multiple lessons over the testing period, resulting in 56 unique survey responses. The overall SUS for the crawl lessons included 44 reported responses with an average SUS score of 70.7 (STD=19.48) as shown in Table 7. This indicated that users thought it is a good system.

Table 5 SUS results for fourth user evaluation

	Fourth User Evaluation		
Lesson	**SUS**	**SD**	**N**
Airway Tools & Treatment	73.8	19.45	2
Respiration Symptoms &Injuries	68.7	19.25	15
Chest Seal Tools & Treatment	72.5	19.28	13
Chest Seal Skills Practice	70.9	21.79	14
Overall	***70.7***	***19.48***	***44***

The SUS for the Respiration Symptoms and Injuries included 15 responses for this lesson and had an average of 68.7 (STD=19.2). This indicates that the users thought the lesson was good in terms of usability. Users noted many positive aspects of the lesson such as the visual details and audio related to the symptoms of the virtual casualty, and the prompting to embody the process.

The SUS for Chest Seal Tool and Treatment included 13 responses and had an average of 72.5 (STD=19.3). This indicates that the users thought this lesson was good in terms of usability. Users noted that what they liked most about the lesson included the length of the lesson, and the animations demonstrating how the user would have to move, allowing the user to imagine themselves doing the steps for the interventions.

The SUS for Chest Seal Skills Practice included 14 responses for this lesson and had an average of 70.9 (STD=21.8). This indicates that the users thought this lesson was good in terms of usability. Among the positive comments for the lesson, users noted that they liked how the hands-on practice was integrated with the XR presentation.

The SUS for Airway Tools & Treatment included 2 responses for this lesson and had an average of 73.8 (STD=21.8). This indicates that the users thought this lesson was also good in terms of usability.

The SUS for Respiration Assessment Practice 2 included 14 responses for this lesson and had an average of 59.2 (STD=15.9). As a run, scenario lesson, the format was new and different from the previous lessons. The responses from this lesson showed more negative responses for the positively coded questions and more positive responses for the negatively coded questions such that responses were grouped around the central responses (Disagree, Neutral, and Agree). Based on user feedback from open-ended responses, the users were frustrated by the alignment of the translucent virtual casualty and the minimal instructions for a run-level lesson which was different from the lessons they had tested leading up to the hands-on practice lesson. In prior lessons, the virtual casualty was opaque and alignment differences were not as pronounced. The results indicated that the virtual symptoms would have to be represented in a different way so users could focus on learning the information rather than being distracted by the representation.

Most system features were well understood, as indicated by scores of 3.4 and higher and an average of 3.7 (see Table 2). The users who had the most difficulty with the placement of the model were those using the Respiration Symptoms and Injuries lesson (the first lesson that the users were exposed to) and the Respiration Assessment Practice 2 lesson where the alignment between the virtual casualty and the manikin was unclear for some users. This finding supports the general trends from the SUS where the users found that the system was good since they were learning the system and the content for the first time.

Heuristic Evaluations

The expert heuristic evaluations addressed usability issues focused on system interactions. Using the Derby Dozen framework, each expert evaluated two lessons over two rounds of review. In the first round, both experts reviewed the Respiration Symptoms & Injuries lesson because it included important content that they would need to understand the other lessons they were reviewing. The seven lessons reviewed are shown in Table 6.

The Derby Dozen checklist targets an evaluation of the system in terms of usability. The checklist was modified to cover the following 8 dimensions: Instructions, Organization & Simplification, Consistency & Flexibility, Integration of Physical & Virtual Worlds, User Interaction, Comfort, Feedback to the User, and Intuitiveness of Virtual Elements.

Table 6 Lessons reviewed as part of heuristic evaluations

Lesson	Expert 1	Expert 2
Chest Seal Skills Practice		X
Respiration Symptoms & Injuries	X	X
Chest Seal Tools & Treatment	X	
Hemostatic Dressing Tools & Treatment		X

Massive Hemorrhage Scenario 1		X
Respiration Assessment Practice 2	X	
Airway Tools & Treatment	X	

The first round of expert evaluations identified areas where additional results from the STA could be applied, specifically where audio and visual cues could create a greater connection to the user's understanding of the situation. By incorporating more cues, the trainee would be able to learn symptoms and cues that would help them decide on the correct intervention. The evaluation also identified areas where activities could be better explained to provide the user more guidance and support throughout the learning process.

The second round of expert evaluations identified areas where the lesson overly relied on text to convey information to the user. Although there were audio and visual cues associated with steps in the lessons, the text would benefit from being streamlined and divided into smaller segments of information. The evaluators found the animations helpful in focusing the user's attention on the 3D space and recommended added more throughout the lessons. Evaluators recommended more visual prompting to help guide the user through the learning process rather than adding more text.

DISCUSSION

As training technologies emerge with low cost, immersive solutions, the need for evaluating the usability of the platform is crucial to determining the effectiveness of the solution for improved performance. In keeping with the TEE framework, the usability results inform the optimization of human-system interaction to develop a training platform built to address the training needs identified. Incorporating recommended changes from usability testing and heuristic evaluations means the application improvements minimize any usability-related effects on performance measures. The following findings were identified as generalizable principles to consider when developing mobile XR applications reaching Kirkpatrick's level 4, impact of the training system on the mission.

Lesson Development

An important finding from the first user evaluation focused on the lesson presentation. The scenario-based lesson was engaging, quick, and focused on one intervention. The second lesson was long, included introductory information on various interventions associated with Massive Hemorrhage treatment. The feedback made it clear that for better engagement and learning, lessons would need to be a shorter format and focused on a theme. For the third user evaluation, the lesson that had been too long was revised to focus on one intervention. New lessons were shorter, included more audio and visual cues, and animations to improve engagement with the learning materials. It is recommended that utilizing a deliberate schema following Bloom's taxonomy for lesson organization should be instantiated when using mobile XR training applications.

Understanding XR

The usability feedback indicated that users need guidance to learn how to use the mobile XR application. The use of XR in a mobile application is different from a head-mounted display which usually offers full immersion. With a mobile application, the user must view content through a phone or tablet to experience the virtual environment. While a mobile application can also create a multi-sensory, multi-layered experience, combining virtual audio and visual cues with reality, the interactions are all through the mobile device which requires the user to understand how to navigate an application differently than normal use. As mobile applications increasingly use XR technology, users will become more familiar with the interactions and optimize the immersive experience that mobile XR affords. Information from the first two user evaluations indicated that users were interested in using the application but were unsure about how to interact with the XR activities and assessments. Based on feedback that users wanted to interact more with the 3D content, the lessons tested at the fourth user evaluation has an increased number of XR interactions. The lower SUS and system feature scores highlight that users needed guidance to understand that the activities require them to embody the activities, getting closer to the casualty to hear and see cues rather than pinching to zoom as they would with other applications. A resulting recommendation is that a short quick start lesson to understand the XR interactions on a mobile device should be included to improve performance-enhancing training.

Hands-on Challenge
A challenge that exists for mobile devices includes how to address and incorporate hands-on skills practice when using a hand-held device. Due to the mobile nature of the training platform, the user must place the device on the ground before practicing hands-on skills. The solution that was tested at the third user evaluation was using a translucent virtual casualty that would overlay symptoms over a manikin or live role player. However, the users found this solution frustrating because of the alignment with real manikins which varied in size. Since the manikin or live role player can be different sizes, the symptoms displayed on the screen were not always in the correct places. Most importantly, the user lost access to the information when they put the phone down to do the hands-on activity. This feedback is crucial to iterating on the solution to promote learning and minimize disengagement because of annoyances with the function of the system. It is recommended that an overhead camera view should be developed with automated realignment of the virtual environment so that content can be viewed while hands on interactions are completed.

Maximizing Audio and Visual Cues
One of the advantages of using XR is the ability to approach learning materials from multiple angles. For example, user evaluations and expert heuristic evaluations both recognized the value added through detailed visual and audio cues that are difficult to relay through the classroom-based teaching lectures. Icons and text in the 3D space mapped back to symptoms or information teach users what to look for in real situation. This direct connection to learning content and the casualty and surrounding environment, immerses the user and reinforces the cues that help the user in the decision-making process. In addition, the training platform offered animations demonstrating hands-on activities that end users could view close-up and from different angles to better understand how they would perform those same interventions. This leads to better cognitive, psychomotor, and affective understanding of the content being learned. In the final performance outcome, if the end user performs the intervention quickly and with greater accuracy, an indicator of affective learning, it would be due to the additional cues provided by the training system. It is recommended that mobile XR systems incorporate audio and visual cues to maximize the multisensory experience afforded by XR.

CONCLUSION

As mobile XR applications proliferate to address training needs, TEEs must examine the cognitive, psychomotor, and affective learning provided by the platform in addition to the performance outcomes on the target measures. By using a life-cycle approach to developing a TEE, the training system is developed alongside the TEE which allows for the review of costs, benefits, and risks of the training system relative to the stated goals of the system. The process of usability testing the training system early and often in the development cycle ensures that the system is considered usable to the end user. Additionally, it is recommended that a preliminary TEE be conducted to further isolate usability issues from effectiveness measures. The findings also contribute to the analysis of learning using the system, determining how well the system provides cognitive, psychomotor, and affective learning. By including the levels of learning, the TEE can effectively demonstrate learning and in what ways the system is helping the end users learn.

REFERENCES

Brooke, J. (1986). System usability scale (SUS): a quick-and-dirty method of system evaluation user information. *Reading, UK: Digital equipment co ltd*, *43*, 1-7.

Cohn, J. V., Stanney, K. M., Milham, L. M., Jones, D. L., Hale, K. S., Darken, R. P., & Sullivan, J. A. (2007). Training evaluation of virtual environments. In E. L. Baker, J. Dickieson, W. Wulfeck, & H. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 81-105).

Derby, J. (2023a). AR Heuristics Tool: 12 Usability Heuristics for Augmented Reality and Mixed Reality. Augmented Reality for Enterprise Alliance. https://thearea.org/free-ar-heuristics-tool/

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

Fitzhugh, D. C. (2015). *Technology Solutions for Combat Casualty Care*. Newsletter HPT&B, 3, 3-5.

Kirkpatrick, D. L. (1959). *Evaluating training programs* (2nd ed.). San Francisco, CA: Berrett Koehler.

Merlo, J.L. (2018). Military operations: Humans not machines make the difference. In R. Steinberg, S. Kornguth, and M.D. Matthews (Eds.), Neurocognitive and physiological factors during high-tempo operations (pp. 195-207). London: Routledge.

Moralez, L., Hughes, C., Archer, J., Miller, C., Stanney, K. M., Frye, D., Karluk, F., & Vazquez, R. (2021, December 29). *Usability Evaluation of Augmented Reality Training for Battlefield Care Tasks*. Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, FL. https://www.xcdsystem.com/iitsec/proceedings/index.cfm?Year=2021&AbID=96944&CID=862#View

Sauro, J. (2011). Measuring System Usability with the System Usability Scale (SUS). Retrieved from https://measuringu.com/sus/

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