

# Mixed Reality Training of Military Tasks: Comparison of Two Approaches Through Reactions from Subject Matter Experts

Roberto Champney<sup>1</sup>(✉), Julie N. Salcedo<sup>1</sup>, Stephanie J. Lackey<sup>1</sup>,  
Stephen Serge<sup>2</sup>, and Michelle Sinagra<sup>1</sup>

<sup>1</sup> Design Interactive, Inc., Orlando, USA  
{roberto, julie.salcedo, stephanie.lackey,  
michelle.sinagra}@designinteractive.net

<sup>2</sup> Institute for Simulation and Training,  
University of Central Florida, Orlando, USA  
sserge@ist.ucf.edu

**Abstract.** This paper discusses a training-based comparison of two mixed reality military trainers utilizing simulation elements that are categorized on different areas of the virtuality continuum. The comparison encompassed exposing subject matter experts (SMEs) to the training systems. Independent groups of SMEs interacted with each system through conducting expert system evaluations. Independent groups of military officers experienced each system for call for fire/close air support training. Following these exposures, participants were queried on the constructs of simulator sickness, training utility, simulator fidelity, usability, and immersion. The results are contrasted and discussed. The outcomes of this comparison serve to promote discussion among the scientific community concerning the training tradeoffs affected by the virtuality continuum.

**Keywords:** Augmented reality · Training · Learning · Immersive training · Virtual reality · Wearable technology · Mixed reality · Training systems · Augmented virtuality · Simulation-based training · Joint forward observer · Call for fire · Close air support · Simulator fidelity

## 1 Introduction

The use of mixed reality in simulation-based training has gained popularity due to its ability to blend real and virtual elements of an experience which may increase the effectiveness of instruction. One particular aspect contributing to the value of mixed reality is its ability to present a compelling contextual experience without the complete artificiality of virtual reality or the risks or costs involved with a completely real experience. Within the realm of mixed reality, there is a continuum of virtuality that ranges from a fully real environment to a completely virtual environment (Milgram et al. 1994). A variety of training systems spanning across the mixed reality spectrum have been prototyped and developed for military fire support task domains including Call for Fire and Close Air Support (CFF/CAS). However, there is presently limited

empirical evidence to inform how much exposure and what type of “reality” or “virtuality” is necessary for training purposes. Current approaches to address this issue primarily focus on optimizing the amount of fidelity (e.g., the amount of realism) in the simulation (e.g., Milham et al. 2008a,b). These methods seek to characterize the necessary sensory, psychological, and functional cues necessary to successfully execute a task in an operational environment in order to prescribe what elements must be provided within a simulation trainer. Given that it is unlikely to replicate 100 % of the cues, one may optimize the amount of fidelity provided based on the criticality of these cues for successful task execution. Even with such methods, decision-makers must determine how to implement the cues and select an appropriate intervention along the virtuality continuum. Stakeholders and decision-makers may weigh risk and cost variables when deciding which elements should be virtual (e.g., military ordnance, aircraft / vehicles, damage), but other aspects may not be as straight forward to decide (e.g., on site location versus virtual location). While the training capabilities of systems may be similar, other aspects of the user experience may be impacted by the approaches taken to provide the simulation. This paper seeks to explore some of these factors by comparing the impressions of subject matter experts after interacting with these systems during operational training exercises.

## 2 Background

The emergence of virtual reality (VR), and later mixed reality (MR), has sparked the development of a broad range of training systems given its ability to provide a viable pedagogical venue for the training of a variety of domains. In particular there seems to be value in merging elements of the physical real-world environment with virtual computer generated imagery to provide an “authentic” participatory experience which can increase the effectiveness of instruction, while improving trainee attention, engagement, and motivation (Kamarainen et al. 2013); particularly in supporting situated and constructivist learning theories involving authentic inquiry and active observation (Dunleavy and Dede 2014). The current study sought to examine the learning reactions and user experience reactions from subject matter experts (SME) from distinct training approaches to the same operational domain. This study consisted of a formative evaluation of two Call for Fire (CFF)/Close Air Support (CAS) simulators. Model A was a portable-outdoor capability augmented reality (AR) system incorporating a head-mounted video-see through display, accompanying backpack hardware and fully functional simulated Binocular tool. Model B was an augmented virtuality (AV) indoor system incorporating an optical-see through display inside a darkened enclosure and representative props for Binoculars and Compass tools. Both systems utilized a real map, pen and notepad for manual tasks within the training. Of interest in the different approaches was the perceived usability of the systems, the presence and magnitude of any simulation sickness, and the impact on immersion while training with the systems.

## 2.1 Operational Domain

Ground Warfighters often employ artillery or mortars in support of missions. A CFF is the process by which a request is made to execute an attack on a target (U.S. Army 1991a,b; Stensrud et al. 2013). These requests are usually initiated by an expert Warfighter (e.g., Joint Fires Observer, Forward Observer, Joint Tactical Air Controller) who is the communication link between Warfighters in the field and those providing the attack (e.g., artillery, mortars). This observer is usually located away from the fight but in a position enabling visual access to locate and identify targets and the effects of attacks on those targets. As part of this process the observer must identify, determine the location of the target, and develop an attack plan while exercising careful decision-making given the dynamic nature of the environment and the likelihood of friendly forces nearby. The attack itself is executed through a series of communications between the observer and a Fire Direction Center (FDC). The communications determine the availability of assets to support the requested attack and relay the command to attack to the firing units (e.g., artillery, mortars). Given the complexity and risks involved with the execution of the attacks in this domain, training for CFF is multi-faceted. Traditional training for this domain involves classroom training, followed by practical exercises, simulation and finally live fire exercises to demonstrate proficiency in knowledge and skills (US Army 2013).

## 2.2 Mixed Reality Continuum

To better understand the differences offered by the two systems evaluated in this study it is necessary to review the continuum of mixed reality. Conceptually one could define mixed reality as a continuum between the real world and a completely virtual world. Milgram and Kishino (1994) proposed such a continuum in their Taxonomy of Mixed Reality Visual Displays (see Fig. 1). While this represented visual displays the same continuum could be utilized to represent the various types of simulator fidelity (physical, functional and psychological). Augmented Reality (AR) which is utilized in Model A lies closer to the real world given the use of superimposed simulation onto a largely real environment. Augmented Virtuality (AV) which is utilized in Model B lies closer to the virtual world given the use of a primarily virtual/synthetic environment augmented by select real elements.

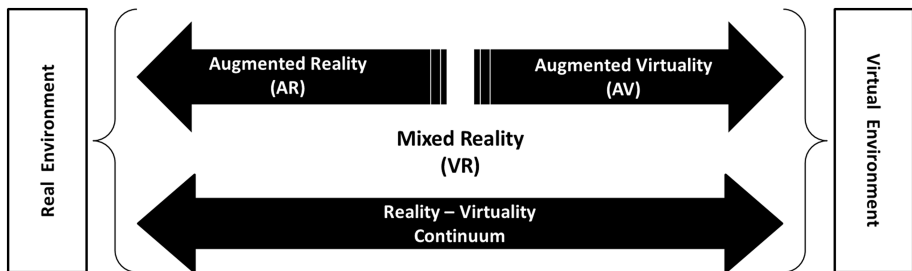


Fig. 1. Taxonomy of mixed reality visual displays (Adapted from Milgram and Kishino 1994)

### 3 Participants

The study was divided into two separate data collection events—one for each system. For Model A – AR, the participants were comprised of five (5) U.S. Marines with prior training and field experience in the CFF/CAS domain. Participants had an average of 8.15 (SD: 2.36) number of years of military service and reported their current role as Forward Observer, EWS Student, or Joint Tactical Air Controller (JTAC). All participants were male and ranged in age from 24 to 38 years old ( $M = 31$ ,  $SD = 6.8$ ).

For Model B – AV, the participants were comprised of three males with prior Reserves training and field experience in the CAS domain. Participants had an average of 10.25 (SD: 2.84) of experience military service and reported their current role as a Joint Terminal Attack Controller (JTAC), Tactical Air Control Party (TACP) specialist, and Forward Air Controller for both the U.S. Marine Corps (USMC) and the U.S. Air Force (USAF). All participants were make and ranged in age from 27–32 years old ( $M = 29.67$ ;  $SD = 2.52$ ).

### 4 Materials

The following tools were used to gather data during the study.

*System Usability Scale (SUS)*: The systems' global usability was evaluated using the SUS (Brooke 1996), a 10-item Likert scale anchored by 1-Strongly Disagree and 5-Strongly Agree (providing a total score ranging from 0–100).

*Simulator Sickness Questionnaire (SSQ)*. The SSQ (Kennedy et al. 1993) was used to assess the incidence and severity of adverse symptoms associated with using the training simulation systems. The SSQ consists of a checklist of 16 symptoms, each of which is related in terms of degree of severity (none, slight, moderate, severe), with the highest possible total score (most severe) being 300. A global score reflecting the overall discomfort level known as the Total Severity (TS) score is obtained through a weighted scoring procedure; three subscales representing dimensions of simulator sickness were also calculated (i.e., Nausea [N], Oculomotor Disturbances [O], and Disorientation [D]).

*Immersion Questionnaire*: A subset of items selected from Jennett et al. (2008) was used to assess the level of cognitive absorption and flow (i.e., the sense of “loosing oneself” in the simulation). Participants responded to the questionnaire using a five-item Likert scale anchored by 1-Strongly Disagree and 5-Strongly Agree.

### 5 Method

User experience researchers and developers define the user experience as “a consequence of a user’s internal state, the characteristics of the designed system, and the context within which the interaction occurs” (Hassenzahl and Tractinsky 2006;

Law et al. 2009). Under this definition, it is recommended that user experience evaluation focus on the impact of the system characteristics and context of the user's psychological state or well-being (Law et al. 2009). In the present evaluation, user experience was evaluated subjectively along several constructs identified to have a potential impact on the effectiveness of simulation-based training platforms including: simulation sickness, usability, and immersion.

The data collection conditions differed slightly across the two systems in order to maintain their operational usecase conditions (i.e., how they would be utilized in a real world application; outdoors versus indoors), yet the methods were kept as similar as possible.

The experimental procedure consisted of pre-exposure, system familiarization, usage, and post-exposure phases. In the pre-exposure phase, participants were provided a brief of the experiment purpose, potential risks and benefits, and experimental tasks. Participants signed the informed consent document if they agreed to participate, and completed the Demographics Questionnaire, as well as a pre-exposure SSQ. Next, in the system familiarization phase, an experimenter assisted the participant with attaching the HMD and then explained the features, tools, and controls to be used during the scenario. The participant was given the opportunity to adjust the HMD for comfort and to gain familiarity with the system's environment and interactions within the system. During the usage phase participants were able to experience a training scenario by observing a target, interacting with simulation tools and observing an attack on a target. Following this, the participant completed the post-exposure questionnaires including: post-exposure SSQ, Subjective Usability Scale, Immersion Measure, and additional questionnaires (these additional questionnaires were varied across the two conditions and thus not included in this manuscript) (Figs. 2 and 3).



**Fig. 2.** Model A – Augmented Reality (AR) system

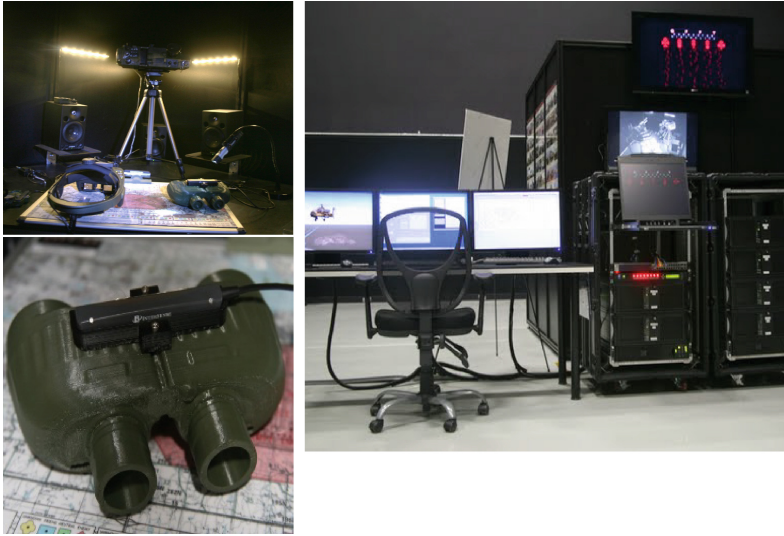


Fig. 3. Model B – Augmented Reality (AR) system

## 6 Results and Discussion

### 6.1 Usability

The SUS was used as a global measure of usability perceived by the participants when interacting with the system. Adequate usability may have training implications as incidences of confusion or frustration with the technology may detract from the learning experience. The results indicated that Model A had an average score of 54 (SD: 17.01) and Model B an average score of 82 (SD: 5) which was significantly higher than Model A. A score of 54 corresponds with an objective measure of “OK” usability (Bangor et al. 2009) and a score of 82 corresponds with “Good-Excellent” usability (see Table 1).

Table 1. Subjective usability scale results

	Model A		Model B		Independent samples T-Test	
	M	SD	M	SD	t	p (2-tailed)
<b>SUS Score</b>	54.00	17.01	82.50	5.00	-2.75	0.033*

Based on responses from the individual questions in the SUS it can be determined that for both Models the detractors from usability were the perceived complexity of setting up the system for use (e.g., “I think I would need the support of a technical person to be able to use this system, Model A M: 2.2, SD: 1.1 and Model B M: 3.67, SD: 0.58,  $t$ : -2.48,  $p$ : 0.048; and “I need to learn a lot of things before I could get going

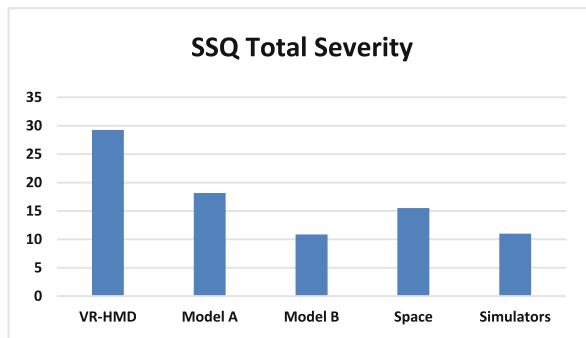
with this system” Model A M: 1.60, SD: 1.34 and Model B M: 3.00, SD: 0,  $t: -2.33$ ,  $p: 0.080$ ). Both systems were prototypes and as such, additional refinement to ensure a high level of usability would be beneficial. Key differences were observed between the two models with regards to specific statements. Specifically, participants believed Model B was significantly more unnecessarily complex than Model A. Yet, participants found Model B (M: 2.6, SD: 0.55) to be better integrated than Model A (M: 4.00, SD, 0.00;  $t: -5.72$ ,  $p: 0.005$ ) (Table 2).

**Table 2.** Subjective usability adjective ratings

Adjective	Adjective rating	
	M	SD
Worst Imaginable	12.5	13.1
Awful	20.3	11.3
Poor	35.7	12.6
OK	50.9	13.8
Good	71.4	11.6
Excellent	85.5	10.4
Best Imaginable	90.9	13.4

## 6.2 Simulator Sickness

Participants interacting with the training systems were assessed for simulator sickness symptoms before and after exposure. It is a common practice in simulator sickness studies to limit exposure to participants demonstrating pre-existing symptoms (i.e., SSQ > 7.48; for example see Champney et al. 2007), however, this was not viable under the current study as it would not be representative of the operational conditions (i.e., Warfighters would need to utilize the system regardless of any pre-existing discomfort such as fatigue, a headache, etc.). While no participants were excluded from



**Fig. 4.** Simulator sickness results in comparison with other similar VR systems (Stanney et al. 1998).

participation based on their pre exposure SSQ score, the data were analyzed for both participants having an SSQ score of 7.48 or less, and all participants regardless of their pre-existing symptoms (Fig. 4).

**Table 3.** SSQ total score before and after training

Model	Pre-exposure		Post-exposure		Paired sample T-Test	
	M	SD	M	SD	<i>t</i>	<i>p</i> (one-tailed)
A	10.15	8.00	18.17	15.83	1.1963	0.2547
B	8.00	13.00	10.87	15.74	-3.85	< 0.001*

\**p* ≤ .025

**Table 4.** Virtual reality stimulus dose (Source: Stanney et al. 2015)

VR stimulus dose	SSQ percentile	SSQ score
Low	25 <sup>th</sup>	< 15.5
Moderate	50 <sup>th</sup>	< 20.1
Medium	75 <sup>th</sup>	< 27.9
High	95 <sup>th</sup>	< 33.3
Extreme	99 <sup>th</sup>	< 53.1

The amount of time participants were in the training systems ranged from 29 to 45 min. Given that the likelihood and intensity of simulation sickness is influenced by exposure time, the amount of exposure to the simulated environment should be taken into account. Ideally the amount of exposure would not produce problematic simulator sickness (i.e., a SSQ Total Severity [TS] score < 20.1 based on Kennedy et al. 2003; Stanney 2001). The results indicated that Model A showed an average post exposure TS SSQ score of 18.17 (SD = 15.63) which is considered a “moderate” dose spectrum as compared to VR systems (see Table 3) and thus even susceptible individuals would be expected to be able to tolerate a single live fire exercise without problematic simulator sickness. Model B had a post-exposure Total Score significantly lower than 20 with a mean of 10.87 which is considered a “low” severity score. While there was a significant difference between Pre-Exposure to Post-Exposure for participants in Model B, this difference was not significant for Model A. There was also no significant difference between Model A and B Post-Exposure TS (*t*: 1.1371; *p*: 0.2611) (Table 4).

To further explore the types of sickness symptoms being experienced immediately following exposure to the simulators, the SSQ profile subscales<sup>1</sup> of Nausea (N), Oculomotor disturbance (O) and Disorientation (D) were assessed (see Table 5). As

<sup>1</sup> Nausea symptoms include general discomfort, increased salivation, sweating, difficulty concentrating, stomach awareness, and burping. Oculomotor symptoms include general discomfort, fatigue, headache, eye strain, difficulty focusing, difficulty concentrating, and blurred vision. Disorientation symptoms include difficulty focusing, nausea, fullness of the head, blurred vision, dizziness with eyes open or closed, and vertigo.



shown in Table 5, Model A had an  $O > N > D$  SSQ profile. This profile is similar to that observed in other military simulators ( $O > N > D$ ) and is different from traditional virtual reality profiles ( $N > O > D$ ) (Kennedy et al. 2003). Model B had a different profile where the Oculomotor subscale had the highest severity score followed by Disorientation and then Nausea. This pattern ( $O > D > N$ ) is different from other military simulators ( $O > N > D$ ) and VE profiles ( $N > O > D$ ) (Kennedy et al. 2003). Given that the oculomotor SSQ subscale score is highest for both systems than those of the Nausea and Disorientation subscales, this indicates that the visual display characteristics of Model A and B system may be responsible for the reported sickness symptoms (i.e., headache, difficulty focusing).

**Table 5.** Simulation sickness profile

Groups	Simulator Profile
Model A	$O > N > D$
Model B	$O > D > N$

While the profile patterns were reversed for N and D subscales, it is challenging to determine the implications of the results of these two subscales. This is because Model A is an outdoor system whose N scale symptomology may be influenced by the local climate or location (i.e., sweating, difficulty concentrating) and was likely driven higher by the operational environment (e.g., hot summer day in direct sunlight). Nonetheless, the fact that oculomotor-related symptoms produced the highest subscale results implies that the training system had a greater impact over the operational environment (whether indoors or outdoors).

### 6.3 Immersion

The results of Immersion assessment showed that in general participants were able to be absorbed by the simulation. This is observed by an average immersion score that is different from the neutral response in the scale (i.e., 3 in a 5 point Likert scale; Model A  $t$ : 2.9397,  $I$ : 0.0260 and Model B  $I$ : 28.9013,  $p$ : 0.0001). The results also indicate that the average immersion rating was significantly higher for Model B than for Model A (see Table 6).

**Table 6.** Immersion rating

	Model A		Model B		Ind. Samples T-Test	
	M	SD	M	SD	$t$	$p$ (2-tailed)
<b>Immersion</b>	3.40	0.36	4.19	0.27	-2.76	0.04*

## 7 Discussion

The usability results (e.g., relatively high SUS scores) to indicate generally good acceptance by participants. Model B ratings were higher than those of Model A which may be attributed to what participants observed in terms of technical glitches or

complexities experienced while interacting with the system. The fact that Model A is a wearable system makes the nature of the prototype more evident as cables and glitches are more directly experienced by the user. In contrast, Model B seemed more robust and glitches were less apparent to the participant, but could be maintained through the operator/instructor interface (away from the participant). This is evidenced by participants' rating of unnecessary complexity and integration of system functions, in both of which Model B was rated better. Other relatively low ratings seemed to be related to intimidation by the technology. Participants' lowest ratings for both systems were recorded for the question "I think I would need the support of a technical person to be able to use this system". Given that participants did not have to setup or activate the system itself but rather just use the system to execute CFF tasks, it is believed these ratings stem from technical intimidation given the highly technical setup required for use of the systems based on what they observed from the operators.

With regards to Simulator Sickness, both systems rated very well given the amount of severity of the reported symptoms. Both post-exposure average SSQ Total Scores were below the 20.1 threshold for Moderate sickness. Given that the likelihood and intensity of simulation sickness is tied to the amount of exposure (Nelson et al. 2000) it is necessary to ensure that exposure times are related to the expected amount of time in an operational training exercise. The amount of time used in this study was not based on training expectations and thus, care should be taken when making inferences regarding simulation sickness expectations under training conditions. Similarly, care should be taken with participants' after-exposure handling given the unknown timeline of symptom progression. Past studies have found symptoms and after-effects remain for one hour or longer (Champney et al. 2007). In particular oculomotor disruption (e.g., eyestrain, inability to focus) should be further studied given the operational domain under which Warfighters operate and the circumstances in which they depend on visual capabilities to operate safely and effectively.

In general, participants felt more immersed in Model B than in the Model A system. Although the observations taken showed that all participants were able to successfully execute and train on the operational tasks, participants indicated that Model B produced a more immersive experience. These findings are not surprising given the approach used. Model B produced an experience where participants were moderately shielded from the surrounding environment and "transported" to an operational virtual world. In contrast, Model A moderately relied on the real environment onto which virtual content was superimposed through a limited field of view apparatus. The differences between the real and virtual world were much more evident in Model A given this approach where real stimuli (e.g., ordnance, aircraft, people) and events surrounded the participant. For instance, while the visual detail of virtual artifacts in the environment were of high quality they possessed artificial characteristics compared to the real objects. The same could be said of sound and other physical elements of a higher intensity (e.g., real explosive ordnance versus artificial ordnance through wearable speaker). Model B in contrast provided a more isolated experience which may have contributed to the higher level of immersion. During the scenario, the participant interacted with the system within the 7' × 7' × 7' enclosure. Although the participant was separated from the SME instructor/role player by a partition, he still interacted with the SME via two-way radio communication.

**Acknowledgements.** This material is based upon work supported in part by the Office of Naval Research (ONR) under contract N00014-12-C-0216 and the Army Research Laboratory (ARL) under contract W911QX-13-C-0052. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views or the endorsement of the ONR or ARL.

## References

- Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: adding an adjective rating scale. *J. Usability* **4**(3), 114–123 (2009)
- Brooke, J.: SUS: a “quick and dirty” usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) *Usability Evaluation in Industry*. Taylor and Francis, London (1996)
- Champney, R.K., Stanney, K.M., Kennedy, R.S., Hash, P., Malone, L., Compton, D.: Recovery from virtual environment exposure: expected time-course of symptoms and potential readaptation strategies. *Hum. Factors* **3**(49), 491–506 (2007)
- Dunleavy, M., Dede, C.: Augmented reality teaching and learning. In: Spector, J.M., Merrill, M. D., Elen, J., Bishop, M.J. (eds.) *The Handbook of Research for Educational Communications and Technology*, 4th edn, pp. 735–745. Springer, New York (2014)
- Jennett, C., Cox, A.L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A.: Measuring and defining the experience of immersion in games. *Int. J. Hum. Comput. Stud.* **66**(9), 641–661 (2008)
- Johnson, L., Kamarainen, A.M., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M.S., Dede, C.: EcoMOBILE: Integrating augmented reality and probeware with environmental education field trips. *Comput. Educ.* **68**, 545–556 (2013)
- Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* **3**(3), 203–220 (1993)
- Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* **E77-D**(12), 1321–1329 (1994)
- Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented Reality: A class of displays on the reality-virtuality continuum. In: *SPIE 2351 Telem manipulator and Telepresence Technologies*, p. 282 (1994)
- Nelson, W.T., Roe, M.M., Bolia, R.S., Morley, R.M.: Assessing simulator sickness in a see-through HMD: Effects of time delay, time on task, and task complexity (ADA: 430344). Air Force Research Laboratory, Wright-Patterson Air Force Base (2000)
- Milham, L.M., Carroll, M.B., Jones, D.L., Dean, S.E., Chang, D.: Cue fidelity evaluation: a requirements-driven approach to training effectiveness evaluation. In: *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (IITSEC) Annual Meeting*, Orlando, FL (2008a)
- Milham, L.M., Carroll, M.B., Stanney, K.M., Becker, W.: Training requirements analysis. In: Schmorow, D., Cohn, J., Nicholson, D. (eds.) *The Handbook of Virtual Environment Training: Understanding, Predicting and Implementing Effective Training Solutions for Accelerated and Experiential Learning*. Ashgate Publishing, Aldershot (2008b)
- Stanney, K.M., et al.: Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. *Int. J. Hum. Comput. Interact.* **10**(2), 135–187 (1998). Report sponsored by the Life Sciences Division at NASA Headquarters

- Stanney, K.M., Kennedy, R.S., Hale, K.S.: Virtual environment usage protocols. In: Hale, K.S., Stanney, K.M. (eds) *Handbook of Virtual Environments: Design, Implementation and Applications*, pp. 797–810 (2015)
- U.S. Army: FM 6-30: Tactics, Techniques, and Procedures for Observed Fire. Department of the Army, Washington (1991a)
- U.S. Army: FM 6-30: Tactics, Techniques, and Procedures for Observed Fire, U.S. Army, Washington (1991b)
- Stensrund, B., Fragomeni, G., Garrity, P.: Autonomy requirements for virtual JFO training. *Proceedings in Interservice/Industry Training, Simulation and Education Conference*, Orlando (2013)
- U.S. Army: JCAS Memorandum of Agreement: Joint Fires Observer. U.S. Army, Washington (2013)