Effectiveness of Augmented Reality & Augmented Virtuality

J D Fletcher, James Belanich, Frank Moses, Ashley Fehr
Institute for Defense Analyses
Alexandria, VA 22311
fletcher@ida.org, jbelanic@ida.org, fmoses@ida.org, afehr@ida.org

Jason Moss
U.S. Army Research Laboratory
Human Research and Engineering Directorate (ARL-HRED)
Orlando, Florida
jason.d.moss11.civ@mail.mil

ABSTRACT

A 2015 analysis of published AR/AV research identified 22 reports that examined and reported empirical data concerning the effectiveness of Augmented Reality (AR) & Augmented Virtuality (AV) for education, training, and performance aiding. Current assessments focused on testing the technology itself (e.g., safety; specifications; capabilities) are numerous, impressive, and essential. However, assessments with objective, empirical data about AR/AV effectiveness for training and performance are comparatively rare. Most reports on effectiveness comparing AR/AV to alternatives suggest that AR/AV reduces time (effect size = 0.52 and errors (effect size = 0.81 in performing skilled tasks, increases the amount learned (effect size = 0.44), produces learning that is more resistant to decay (effect size = 0.71), is preferred to other approaches (effect size = 0.81), and increases immersion ("flow") during learning (effect size = 0.67). Other findings of interest were reported from individual studies. However simulator sickness remains a problem. Recommendations from the analysis included: Increased emphasis on empirical assessment of AR/AV effectiveness compared to other systems; research on the development of systems using machine intelligence combined with AR/AV; research on adapting AR/AV to individual differences; exploiting the potential of AR/AV to enhance success in performing military operations in combat service support (e.g., maintenance and repair) activities; cost-effectiveness assessment AR/AV in both training and performance aiding; and continuing review of costs and advances in commercial AR/AV technology.

ABOUT THE AUTHORS

J D Fletcher, James Belanich, Frank Moses, Ashley Fehr
The above authors are all members of the research staff in the Science and Technology Division of the Institute for Defense Analyses
Jason Moss

The above author is a member of the research staff in the Human Research and Engineering Directorate (ARL-HRED) of the U.S. Army Research Laboratory.

Effectiveness of Augmented Reality & Augmented Virtuality

J D Fletcher, James Belanich, Frank Moses, Ashley Fehr
Institute for Defense Analyses

Jason Moss
U.S. Army Research Laboratory
Human Research and Engineering Directorate (ARL-HRED)

AUGMENTED REALITY AND AUGMENTED VIRTUALITY

Augmented reality (AR) and augmented virtuality (AV) are regularly represented as two areas within a continuum of mixed reality (MR), as described by Milgram & Kishino (1994). The continuum spans from a completely real environment to a completely virtual environment, with AR defined as the integration of a real world environment with some virtual information and AV defined as primarily a virtual environment augmented in real time with real world objects and actions. Common examples of AR include the first-down marker that is inserted into TV broadcasts of football and heads-up displays for fighter pilots where information like altitude, heading, and speed are displayed onto their visor (Azuma, 1997; Azuma et al., 2001). Complementarily, AV is identified as primarily a virtual environment that is augmented with some real world objects and actions. Examples of AV would include: a fully-occluded HMD in which the user's perspective of the virtual environment changes in accord with the user's movement and is augmented by real objects and computer-generated visual and audio cues (Caballero et al., 2010); or a system where a user moves through a virtual shoot-house using a realistic weapon for training.

The Milgram and Kishino (1994) MR continuum was based on visual displays and presents MR as a single continuum from real to virtual. However, Milgram and Kishino were assessing only visual displays, while the continuum has been used to describe complete systems that include other dimensions, such as the use of real/virtual audio or other sensory stimulation and different tracking methods (Zhou, Duh, & Billinghurst, 2008). Even for visual displays, the type of display -- a computer monitor, see-through display, head-mounted video display (i.e., HMD), mobile/handheld device, or 3-D video display -- has differing influences on how the mix of real/virtual entities are perceived (Stevens, Kincaid, & Sottilare, 2015). Stevens, et al. (2015) suggest that the type of display may influence how well it performs particular functions, specifically training.

While the MR continuum is neatly shown as a simple continuum, it is not exactly clear where a particular system may fall along that continuum because there is no clear quantifiable

measure of how much real and how much virtual experience is presented to a user, the ratio of real to virtual is not always clear. That problem is amplified with increasingly complex systems include visual, auditory, and haptic senses. Likewise, factors like tracking the user and real world objects along with methods for integrating the real and with the virtual are not considered. This may lead to debates of where a system lies along the continuum. Some of this confusion can be addressed if the components of AR/AV systems are categorized in functional terms:

1) *User perception* - how the user perceives both the real and virtual entities. A user can perceive the real and virtual world through many senses, with vision, hearing, and touch being the most common. Olfaction and taste could also be included in this function, but have not been used extensively and are not addressed in this review. Overwhelmingly visual displays are the focus of AR/AV developers, with sound presentation also receiving considerable attention (Livingston, et al., 2013b). The video screen could display a completely virtual image (VR) or mostly virtual with a little reality included (AV) or most of the screen could incorporate a live video feed of the real world in front of the user with a little virtual components inserted (AR). Depending on the ratio of real versus virtual, the display could be considered either AR or AV. Similarly, there are multiple methods for presenting auditory information to a user in an AR/AV environment (Vörlander & Shinn-Cunningham, 2014). These include the use of a single speaker from a tablet/smartphone, binaural headphones, or multiple speakers. With a single speaker, no spatially oriented information is provided. If the task requires the user to locate the source of the sound or orient towards the sound, at least stereo (binaural) headphones are needed.

2) *Tracking the user's actions* - determining the user's head orientation, body position, and movement in space. Tracking a user can include up to six degrees of freedom, which consist of 3 linear and 3 rotational components (Stewart, 1965). The three linear movements are changes in position/location in the x, y, and z planes, which could be described for a person in an AR/VR environment as walking forwards/backwards, side stepping to the left or right, and going up/down. The three rotational movements are changes in orientation (i.e., directional angle), which could be described for a person in a AR/VR environment as moving the head to look up/down, turning their head to look to the left or right, and tilting their head to so that their eyes were at different heights. The position and motion of a user's head is critical to an AR/AV system because where the head is determines much of what must be displayed to a user (You, Neumann, &Azuma, 1999).

3) *Tracking objects/locations* – determining the position and orientation of real and virtual entities and locations in the environment. The same six degrees of freedom for position and motion relate to objects in the AR/AV environment. Consider for example, training a user in an AR/AV environment to remove a particular screw as part of a vehicle maintenance task. Placing and positioning a virtual label that highlights the screw to be removed is critical to identifying it.

4) *Mixing real and virtual* - software and computer processing to fuse real and virtual entities into a single mixed reality. This requires integrating input from multiple sensors, updating status of the MR environment, and rendering and presenting the appropriate

stimulation to the user. In combat vehicle training this could include GPS and visual tracking methods to determine the location and orientation of the vehicle, a database to understand the surrounding terrain, the human interface components that detect the user's actions, and AR rendering and scenario manager components to bring the real and virtual components together for display on the vehicles monitors.

The goal, intent, and expectation of a user will influence how the mixing of real and virtual information from an AR/AV system (Kruijff, Swan, & Feiner, 2010) is perceived. For example, if a user expects particular real and virtual humans/objects to be about the same size, then the expectation is that they are the same distance away if the retinal image size of each is similar. However, if the retinal image size is different but the user expects the objects to be similarly sized (even though they may not be) there may be a misperception of distance (Kalkofen, et al., 2011). Similarly, misperceptions may arise when visual and auditory cues either link or don't link appropriately (temporally or spatially) based on the expectation of the user as when a person's mouth moves with mistimed audio (Vörlander & Shinn-Cunningham, 2014).

ANALYSIS OF PUBLISHED RESEARCH

AR/AV technology is undergoing rapid advancements and improvements. Empirical data on its effectiveness in training and performance aiding is sparse and scattered across many disciplines and communities of research. Understandably, published training effectiveness research is lagging behind the newest technology developments. However, presently available data and reviews indicate considerable promise. Data summarized in this section were identified and collected using meta-analytic processes to collect and report statistical and practical significance.

The review of published literature focused on articles for which document search engines found research mentioning "evaluation" or "assessment" along with "augmented reality", "augmented virtuality", or "mixed reality" -- or approximate synonyms of these terms. Additional candidate sources were drawn from reference lists in the articles.

The processes used to identify and collect empirical studies of effectiveness studies were the following:

- Use document search engines to locate research materials with one or more of the terms described above in titles, abstracts, and key word lists.
- Scan reference lists in these materials to locate and obtain additional candidate sources, which themselves were then scanned for further sources.
- Select candidate sources that report empirical data on AR/AV effectiveness in training, education, or performance/decision aiding.
- Set aside sources that do not report empirical data. A number of authors were contacted with email requests for additional data. Several provided helpful, much appreciated, responses.

Overall, we found more than a hundred reports, papers, and presentations that mentioned assessment (or approximate synonyms) of AR/AV or mixed reality. A growing body of excellent research on ergonomic design for AR/AV technical performance was found to guide the development and design of the technology itself. This work is relevant and essential, but the impact of its findings on human performance is indirect. Overall we found assessment research to be proportionally sparse compared to research on design and development of the technology.

We found 22 applicable reports. Many of these reports provided data on more than one comparison (e.g., time to perform task, learning, errors, and preference). In all they provided 96 empirical data points (e.g., comparisons) drawn from valid assessments of AR/AV effectiveness in education, training, performance enhancement, decision making, or job aiding. Seventeen of the 22 reports concerned AR (augmentation of real environments and tasks) rather than AV (augmentation of virtual environments). Ten reports concerned performance aiding, and 12 concerned instruction equally divided between education and training. More complete data and detailed information is provided by Belanich, Fletcher, Moses, and Fehr (2016).

Effect sizes are used as much as possible to indicate the practical significance (in contrast to statistical significance) of findings. Effect sizes are reported here as Hedges 'g' (Hedges & Olkin, 1985), which is calculated as a standard deviation from the estimated population mean to indicate the practical significance of the finding after adjusting for sample size. Effect size can be calculated from group sizes, means, and standard deviations and from a variety of inferential statistics (e.g., t-scores, F-ratios, Chi-square). Table 1 suggests an interpretation of effect size magnitude.

Table 1. Overview of Effect Size

Effect Size	Suggested Interpretation[a]	50th Percentile (Roughly) Raised To …
ES < 0.25	Negligible[b]	60th percentile
0.25 < ES < 0.40	Small	60th–66th percentile
0.40 < ES < 0.60	Moderate	66th–73rd percentile
0.60 < ES < 0.80	Large	73rd–79th percentile
ES > 1.00	Very Large	80th percentile and up
ES > 2.00	Bloom's challenge[c]	98th percentile and up

a Extended from suggestions by Cohen (1988). b What Works Clearinghouse (2010). c Bloom (1984).

Table 2 summarizes findings from all the experimental studies found in the published effectiveness reports.   A majority of the published studies of AR/AV compared to alternatives indicate that it takes less time, increases amount learned and its persistence, and has fewer errors. The alternative approaches varied across studies. Subjectively, users reported that they preferred AR/AV to a number of alternate approaches and that they were more engaged in

their tasks. Column 1 indicates the measure that was assessed for each row. Column 2 shows the average effect sizes for all study findings that were published, that could be calculated or were obtained from authors for particular measures of performance listed in column 1. Column 3 shows the ratio of all comparisons (i.e., N from Column 2), where AR/AV was favored over the alternative method. Column 4 reports average effect sizes from studies where differences between groups (AR/AV compared to alternative) were statistically significant (p < 0.05). At issue is the degree to which effect sizes differ between significant and non-significant outcomes. All but one comparison found effect sizes to be larger for statistically significant cases.   The first four rows of Table 2 are based on objective data. The last three rows are based on subjective impressions/opinions provided by participants in the studies.

Table 2. Summary of published research findings

Measures	Effect Size[a] Findings: Average, Standard Deviation, & N[b]	Effect Sizes Favoring AR/AV	Effect sizes favoring AR/AV in Statistically Significant (p < 0.05) Comparisons[c]
Performance aiding to reduce time performing tasks	AVG = 0.518 SD = 1.192 N = 29	17/29 (59%)	14/19 (74%)
Increased Learning (Education & Training Combined)	AVG = 0.441 SD = 0.458 N = 30	26/30 (87%)	17/19 (89%)
Increased Learning persistence	AVG = 0.708 SD = 0.226 N = 11	11/11 (100%)	7/7 (100%)
Fewer Errors	AVG = 0.808 SD = 0.609 N = 9	9/9 (100%)	8/8 (100%)
Preference for AR/AV	AVG = 0.811 SD = 1.823 N = 8	6/8 (75%)	5/6 (83%)
Increased immersion ("Flow")	AVG = 0.664 SD = 0.443 N = 5	5/5 (100%)	4/4 (100%)
Increased immersion persistence	AVG = 0.785 SD = 0.457 N = 4	4/4 (100%)	4/4 (100 %)

[a]All effect sizes are Hedges g (Hedges & Olkin, 1985); [b]Some published reports included more than one comparison, and each comparison was considered uniquely; [c] Where statistical significance was obtained and effect size could be calculated.

The findings were largely positive. However, there also are studies that contradict the majority. For instance, AR/AV in some studies took more time and the amount learning may be only moderate or small. The following comments apply to the rows of Table 2.

Task Performance Time: Seventeen of twenty-nine (59%) studies of performance/decision aiding comparing AR/AV to an alternative found that time to perform training tasks was reduced through the use of AR assistance with a 'moderate' effect size of 0.518. Of the comparisons where AR/AV were not favored to an alternative, 5 were drawn from a single report concerning collaborative performance (Billinghurst et al. 2003). All five of these observations, which were of 2-person teams rather than individuals, found longer times to perform a task. They found an average effect size of 0.890 (0.647) in favor of alternatives to AR/AV. The remaining 24 studies, in which tasks were performed by an individual, found a 'very large' effect size of 0.812 (1.065). More empirical work is especially needed to clarify the time factor in individual vs. collaborative performance, because so many workforce tasks are performed by teams.

Task Performance Proficiency: Similarly, when task performance time was examined as a product of training (Gavish, et al., 2015; Quarles, et al. 2008; Quarles, et al., 2009), the effect size calculated from the 5 available data points averaged -0.556 (0.814), with the three negative effects obtained from the Quarles studies. Again, more empirical work is needed to clarify time in training task performance, especially in light of the moderate effect size reported in the next.

Learning and error rates: In comparing learning (education or training) under AR/AV conditions with an alternative, 30 comparisons found a moderate effect size of 0.441. Also, average effect size for 11 measures of learning persistence after AR/AV assisted instruction was 'large' (0.708). Additionally, errors in learning were found to be less common in AR/AV applications with 9 comparisons averaging a 'very large' effect size of 0.808. Additional research should investigate underlying cognitive mechanisms that explain these results as well as instructional techniques that might capitalize on them.

Preference and 'Flow': Average effect size for preference of AR/AV-based instruction over other approaches was 'very large' 0.811 (SD = 1.823) -- for the 8 comparisons across the 5 reports that examined it. Two reports (Ibañez, et al., 2014; Zhang, et al., 2014) measured users' experience of Flow, which is characterized by Csikszentmihalyi (1990) as complete absorption in whatever one is doing. It was 'large' with an average effect size of 0.664 (0.443) in 5 comparisons. Further, it persisted with an effect size of 0.785 (0.457) in 4 comparisons that examined it. These outcomes are notable because of the difficulties some learners reported in operating AV/AR systems.

In areas of training or learning overall, reported data includes a wide variety of measures and an equally wide variety of alternatives to AR/AV. Additional work is needed to explain this variety at a cognitive level in order to identify learning objectives for which AR/AV is more effective. At present the data are too sparse to make this assessment. Nonetheless, this review suggests that the use of AR/AV: reduces time to perform manual tasks; increases learning;

reduces errors in task performance; and reduces the decay of learning over time. Finally, the use of AR/AV may increase the sense of immersion or flow, thereby sustaining motivation to learn in the first place and continue learning over time.

Some reports that were included in the review described above also provided additional findings of interest that are relevant to the use of AR/AV, but these findings were not assessed by multiple experiments. These additional findings are the following:

• Baird & Barfield (1999) found that performance time for assembling components for a circuit board to be lower with see-through than monocular displays (g = 0.46) and that users preferred using the see-through rather than monocular displays (g = 0.77).

• Billinghurst, et al. (2003) found that less time was required by 2-member teams to perform an urban design collaborative task using hand-held rather than head-mounted AR displays (g = 0.46) because a wider field of view was readily provided by lowering a hand-held display.

• Dünser, et al. (2012) found both time to complete an outdoor navigation task (g = 0.64) and distance traveled (g = 0.94) to be less for male than female subjects.

• Echeverría, et al. (2012) found greater gains in learning to perform collaborative games with AR/AV (g = 0.91) and in overall learning (g = 0.45) for boys than for girls – girls showed greater learning with a computer mice display than with AR/AV (g = 0.33).

• Henderson & Feiner (2009) found much less time to localize components in armored vehicle maintenance and repair with AR/AV than with an LCD monitor (g = 6.50) or a heads-up display lacking AR/AV registration with items of interest (g = 8.01).

• Henderson & Feiner (2011a) later found AR/AV preferable to use of an Interactive Electronic Technical Manual (IETM) (g = 2.48), but also easier (1.71), more satisfying to (g = 2.52), and more intuitive to use (g =2.09) in performing assembly tasks for a turbo-prop engine.

• Henderson & Feiner (2011b) also found that physical labels on equipment produced slightly better (shorter) performance time than AR did, however, AR produced fewer errors than labeling did.

• In surveys, Ibáñez, et al. (2014) found AR preferable in experimenting with electro-magnetism to World-Wide-Web assistance in terms of concentration (g = 0.86), sense of control (g = 0.67), time sense (g = 0.55), feedback (g = 0.57), autotelic experience (g = 0.86), and skill challenging perception (g = 0.52).

• Macchiarella & Vincenzi (2004) found video AR and interactive AR to be about equally effective in recalling component names and functions -- but both superior to print-based training (g = 0.78) and video training (g = 0.41).

• Martin-Guitérrez, et al. (2010) (not in Appendix B table) found post- minus pre-learning gain scores to be larger for groups receiving 9 hours of AR-based training used to develop spatial visualization and mental rotation abilities (g = 0.73 and g = 0.78, respectively) than groups receiving no extra training.

Summary and Recommendations

This review identified AR/AV developments with potential to support training and performance, but with limited effectiveness data to support this possibility. Overwhelmingly, findings from AR/AV research and assessment concern technology specifications rather than its effectiveness in training and performance aiding.   Findings for performance aiding and training practices point to applications that may substantially improve human performance, but that has been only sporadically verified by objective, empirical evidence. More emphasis on empirical assessment of these applications is needed and recommended.

Recommended next steps to advance the training effectiveness of AR/AV are to:

Increase emphasis and support for empirical assessment of AR/AV effectiveness compared to other systems for improving training and human performance. This research should proceed in step with the rapid enhancement of AR/AV technology itself in both commercial and Defense activities.

Synchronize research on human-system/technology interaction effects with development of AR/AV technology to minimize potential negative effects on the human/user, such as simulator sickness.

Perform research to improve the design of AR/AV-based training and performance-aiding so that it complements capabilities already in place.

Pursue research on the development of systems using machine intelligence combined with AR/AV.

Investigate the potential of AR/AV to support training not only for developing novice and introductory skills but also skill mastery and expertise.

Exploit the potential of AR/AV to enhance success in performing aiding (e.g., maintenance and repair) activities.

Identify the specific variables and components of AR/AV systems that influence their utility for different training and performance objectives to ensure cost effective development of human capabilities and performance.

Establish and maintain continuing review of costs and advances in commercial AR/AV technology to inform decisions involving both material costs (e.g., system development, procurement, modification) and life cycle costs (e.g., personnel, preparation, on-going use, and maintenance). The review should consider return on investment as well as cost.

REFERENCES

Azuma, R. T. (1997). A survey of augmented reality. Presence, 6(4), 355-385.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *Computer Graphics and Applications, IEEE, 21*(6), 34-47.

Baird, K.M., & Barfield, W. (1999). Evaluating the Effectiveness of Augmented Reality and Wearable Computing for a Manual Assembly Task. *Virtual Reality, 4,* 250-259.

Belanich, J., Fletcher, J.D., Moses, F.L., Fehr, A.M.A. (2016) *Assessment of Training Effectiveness with Augmented Reality and Augmented Virtuality* (D-8099). Alexandria, VA: Institute for Defense Analyses.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16.

Billinghurst, M., Belcher, D., Gupta, A., Kiyokawa, K. (2003). Communication behaviors in co-located collaborative AR interfaces*. International Journal of Human-Computer Interactions, 16,* 395-423.

Caballero, M. L., Chang, T. R., Menéndez, M., & Occhialini, V. (2010, September). Behand: augmented virtuality gestural interaction for mobile phones. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (pp. 451-454). ACM.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences: 2nd Edition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. New York: Harper Perennial.

Dünser, A., Billinghurst, M., Wen, J., Lehtinen, V., Nurminen, A. (2012), Exploring the use of handheld AR for outdoor navigation. *Computers &Graphics, 36,* 1084–1095.

Dünser, A., Grasset, R., & Billinghurst, M. (2008). A survey of evaluation techniques used in augmented reality studies (pp. 5-1). Human Interface Technology Laboratory New Zealand.

Echeverría, A., Améstica, M., Gil, F., Nussbaum, M., Barrios, E., Leclerc, S. (2012). Exploring different technological platforms for supporting co-located collaborative games in the classroom. *Computers in Human Behavior 28(4),* 1170-1177.

Gavish, N., Gutierrez, T., Webel, S., Rodríguez, J, Peveri, M., Buckholt, U., & Tecchia, F. (2015). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments (23),* 778-798.

Hedges, L.V., & Olkin, I. (1985). Statistical methods for meta-analysis. New York: Academic Press.

Henderson, S. & Feiner, S. (2011a). Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Transactions on Visualization and Computer Graphics,* 17(10), 1355-1368.

Henderson, S. & Feiner, S. (2011b) Augmented reality in the psychomotor phase of a procedural task. *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR '11)*, Basel, Switzerland, pp. 191-200.

Herling, J., & Broll, W. (2011). Markerless tracking for augmented reality. In Handbook of Augmented Reality (pp. 255-272). Springer New York.

Ibáñez, M. B., Di Serio, A., Villarán, D., & Kloos, C. D. (2014). Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education, 71*, 1-13.

Kalkofen, D., Sandor, C., White, S., & Schmalstieg, D. (2011). Visualization techniques for augmented reality. In B. Furht (ed.), Handbook of Augmented Reality (pp. 65-98). Springer New York.

Korris, J. H., & Garrity, P. (2011). Application of Augmented Virtuality to Close Air Support JTAC Training. Presented at Interservice/Industry Training, Simulation and Education Conference 2011, Orlando, FL.

Kruijff, E., Swan II, J. E., & Feiner, S. (2010, October). Perceptual issues in augmented reality revisited. In ISMAR (Vol. 9, pp. 3-12).

Kulik, J. A., & Fletcher, J. D. (2015). Effectiveness of Intelligent Tutoring Systems A Meta-Analytic Review. *Review of Educational Research*, 0034654315581420.

Livingston, M. A., Rosenblum, L. J., Brown, D. G., Schmidt, G. S., Julier, S. J., Baillot, Y., Swan II, J. E., Ai, Z., & Maassel, P. (2011). Military applications of augmented reality. In *Handbook of Augmented Reality* (pp. 671-706). Springer New York.

Livingston, M. A., Dey, A., Sandor, C., & Thomas, B. H. (2013a). Pursuit of "X-ray vision" for augmented reality (pp. 67-107). Springer New York.

Livingston, M. A., Gabbard, J. L., Swan II, J. E., Sibley, C. M., & Barrow, J. H. (2013b). Basic perception in head-worn augmented reality displays. In *Human Factors in Augmented Reality Environments* (pp. 35-65). Springer New York.

Macchiarella N. D., Vincenzi D. A. (2004) Augmented reality in a learning paradigm for flight aerospace maintenance training. In: *Digital Avionics Systems Conference, vol 1,* (pp 5.D.1–5.1-9). Salt Lake City, UT.

Martin-Guítérrez, J. Saorín, J., Contrero, M., Alcañiz, M., Pérez-López, D., Ortega, M. (2010). Design and validation of an augmented book for spatial abilities development in engineering students. *Computers & Graphics, 34,* 77-91.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, *77*(12), 1321-1329.

Moss, J. D., Austin, J., Salley, J., Coats, J., Williams, K., & Muth, E. R. (2011). The effects of display delay on simulator sickness. Displays, 32(4), 159-168.

Quarles, J., Lanpotang, S., Fischler, I., Fishwick, P., & Lok, B. (2008). A mixed reality approach for merging abstract and concrete knowledge: *Proceedings from IEEE Virtual Reality*. Reno: NV.

Quarles, J., Lanpotang, S., Fischler, I., Fishwick, P., & Lok, B. (2009). Scaffolded learning with mixed reality. *Computers & Graphics, 33,* 34-46.

Stewart, D. (1965). A platform with six degrees of freedom. *Proceedings of the Institution of Mechanical Engineers, 180*(1), 371-386.

Vorländer, M., & Shinn-Cunningham, B. (2014). Virtual Auditory Displays. In Handbook of Virtual Environments: Design, Implementation, and Applications, Second Edition, edited by Kelley S. Hale, and Kay M. Stanney, (pp. 87-114). CRC Press.

What Works Clearinghouse (August 2010). WWC intervention report: High school math, Carnegie Learning curricula and Cognitive Tutor software. Washington DC: US Department of Education.

You, S., Neumann, U., & Azuma, R. (1999). Orientation tracking for outdoor augmented reality registration. *Computer Graphics and Applications*, IEEE, 19(6), 36-42.

Zhang, J., Sung, Y., Hou, H., & Chang, K. (2014). The development and evaluation of an augmented reality-based armillary sphere for astronomical observation. *Computers & Education, 73*, 178-188.

Zhou, F., Duh, H. B. L., & Billinghurst, M. (2008, September). Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (pp. 193-202). IEEE Computer Society.