Translating AR/VR Research into Useable Information for Non-researchers

James Belanich, PhD, Franklin Moses, PhD
Institute for Defense Analyses
Alexandria, Virginia
jbelanic@ida.org, fmoses@ida.org

Emily Fedele, PhD
Institute for Defense Analyses
Alexandria, Virginia
efedele@ida.org

B. Adrian Flowers, PhD, Susannah Hoch, MS
Aptima, Inc.
Woburn, MA
aflowers@aptima.com, shoch@aptima.com

Wink Bennett, PhD
AFRL, 711th Human Performance Wing
Dayton, OH
winston.bennett@us.af.mil

ABSTRACT

Numerous meta-analyses suggest that Augmented Reality (AR) and Virtual Reality (VR) technologies can be effective for training and performance enhancement. However, the broad assertion of AR/VR effectiveness is overly simplistic and clouds important factors influencing the generalizability of findings. To clarify how to use the AR/VR findings, we conducted an initial review of the research literature, organized and simplified knowledge scattered across effectiveness studies, and developed a training effectiveness framework to organize and explain them. The studies reviewed were conducted by those in disparate fields including computer scientists and engineers who focused on technology, educators who focused on instructional methods, and domain specialists (e.g., medical, military, and construction) who focused on implementations in their disciplines. We found many clarity issues, such as lack of: descriptions of users' task experience (26% of studies), AR/VR technology experience (28%), and performance measures (29%). In addition, we found inconsistent use of terms, identification of a system without specifying features used, and limited descriptions of tasks trained. Based on our reviews and discussion sessions with AR/VR professionals, we developed a practitioner-oriented framework to help systematize and bring clarity to the often discrepant research findings. The framework features a common language with five components: What is the technology? What are the skills/tasks trained? How is the technology used for training? Who are the users? What was the outcome? Using this framework, we are populating an online knowledge base that can be queried with specific characteristics that help tease apart the evidence of effectiveness for AR/VR technologies in a nuanced and systematic way. We summarized each study in the knowledge base using the framework's standardized structure to make the empirical findings more coherent, and to help practitioners better translate the research findings into practice.

ABOUT THE AUTHORS

Dr. James Belanich has been a research psychologist at the Institute for Defense Analysis (IDA) since 2012, analyzing human behavior, investigating workforce issues, evaluating STEM (science, technology, engineering, and math) education program, assessing training technology, and modeling intermediate force capabilities. He was also a member of the U.S. Army's Education Advisory Committee (2015-2021). From 2001 to 2012 he was a research psychologist at the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), managing ARI's research program, coordinating institute-wide activities, and conducting research on advanced training methods. He received his Ph.D. in Psychology – Learning Processes from the City University of New York in 2001.

Dr. Franklin (Frank) Moses has been a research psychologist conducting applied research projects for the Department of Defense for more than 45 years. Since 2004, at the Institute for Defense Analysis (IDA), he conducted research about air and ground systems, command and control, intelligence analysis, and personnel, training, and education topics. Prior to 2004, he was a research psychologist at the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) responsible for projects in command, control, communications, computer and intelligence systems. In addition, he had a 4-year assignment to the Office of the Undersecretary of Defense (Acquisition) to review research and engineering projects performed by the three military departments. Dr. Moses received his Ph.D. in applied Experimental Psychology at Tufts University followed by a National Science Foundation postdoctoral fellowship at New York University.

Dr. Emily Fedele is a researcher at The Institute for Defense Analyses. She joined IDA in 2018, and her work focuses on conducting and evaluating behavioral science research on a variety of defense-related topics, including evaluating human effects of intermediate force capabilities, assessing training technology, and providing technical and programmatic evaluation for resilience-based programs. Emily has expertise in research design, experimental methods, and statistical analysis. She received her Ph.D. in Linguistics from the University of Southern California.

Ms. Susannah Hoch is a Research Engineer at Aptima, Inc., with a background in Human-Computer Interaction and Psychology. At Aptima, Ms. Hoch uses her expertise to design and develop tools for a variety of users. Prior to joining Aptima, Ms. Hoch was a software engineer at the Kwajalein Missile Range, where she developed, tested, and provided training for a suite of visualization and analysis tools for the Mission Control Center. Ms. Hoch also worked as a Human-Systems Collaboration Engineer for Draper Laboratory and helped to define their Human-Systems Integration methodology. Ms. Hoch holds a master's degree in Information Science, specializing in Human-Computer Interaction from the University of Michigan and a BA in Psychology from the University of Pennsylvania.

Dr. Winston "Wink" Bennett Jr. received his Ph.D. in Industrial and Organizational Psychology from Texas A&M University. He is the Readiness Product Line Lead for the Airman Systems Directorate at Wright Patterson AFB. He is a Fellow of the Society for Industrial and Organizational Psychology, the American Psychological Association, and the Association for Psychological Science, as well as a Research Fellow at the Air Force Research Laboratory. He is the USAF representative to the NATO Modeling and Simulation Group (NMSG). He is spearheading the Combat Air Forces migration to proficiency-based training and is conducting research related to the integration of live and virtual training and performance environments to improve mission readiness and job proficiency. He served as the Technical Advisor for the Joint Secure LVC Advanced Training Environment Technology Demonstration (SLATE). This was the first fully integrated LVC capability demonstration with live combat-coded aircraft and combat-coded aircraft operational flight program software. He leads research that has developed methods to monitor and routinely assess individual and team performance across live and virtual environments and evaluating game-based approaches for training, work design, and job restructuring. He maintains an active presence in the international research and practice community through his work on various professional committees and his contributions in professional journals and forums including I/ITSEC.

Translating AR/VR Research into Useable Information for Non-researchers

James Belanich, PhD, Franklin Moses, PhD
Institute for Defense Analyses
Alexandria, Virginia
jbelanic@ida.org, fmoses@ida.org

Emily Fedele, PhD
Institute for Defense Analyses
Alexandria, Virginia
efedele@ida.org

B. Adrian Flowers, PhD, Susannah Hoch, MS
Aptima, Inc.
Woburn, MA
aflowers@aptima.com, shoch@aptima.com

Wink Bennett, PhD
AFRL, 711th Human Performance Wing
Dayton, OH
winston.bennett@us.af.mil

INTRODUCTION

Numerous meta-analyses and literature reviews provide evidence that Augmented Reality (AR) and Virtual Reality (VR) technologies can be effective for training and performance enhancement. However, the broad assertion of AR/VR effectiveness is overly simplistic and clouds important factors influencing the generalizability of findings. We can augment and represent reality using many different technologies whose complexities are difficult to organize and understand. In addition, many different kinds of disciplines and publications report on effectiveness studies. In order to make sense of the diverse and diffuse AR/VR research literature, the current work had two main objectives: 1) to develop an analytical framework for AR/VR empirical studies, and 2) to use the framework in ways that meet many needs including those of non-researchers. This report defines and documents a novel framework for making sense of AR/VR.

Background

Literature reviews and meta-analyses are important tools to study commonalities across research projects although both are insufficient for identifying the depth and breadth of AR/VR for training. Literature reviews focus on particular topics of interest to their authors as a subjective way to organize research. They provide multiple perspectives on how to think about AR/VR. Meta-analyses are a systematic way to combine data from multiple studies that emphasize common effects or identify the reasons for variation.

While neither literature reviews or meta-analyses may fully describe the dimensions that influence the use and generalizability of AR/VR effectiveness, meta-analyses in particular helps to identify them. For example, a meta-analysis (Fletcher et al., 2017) comparing AR/VR[1] to training alternatives across 22 reports with empirical data found that performance using AR/VR takes less time (effect size = 0.52), increases amount learned (effect size = 0.44), leads to more persistent learning (effect size = 0.71), and results in fewer errors (effect size = 0.81). In addition, users reported that they preferred AR/VR to alternate approaches and felt more engaged in their tasks. Other analyses identify additional factors about when and how AR/VR may be more or less effective. Kaplan et al. (2020) asked how the use of extended reality (XR, a term referring to any form of AR, VR, and mixed reality) compares to traditional training methods and found XR and traditional training equally effective at enhancing performance. Subsequently, they suggest that XR's main value is in providing training where traditional methods may be dangerous or costly. Another meta-analysis (Batdi & Talan, 2019) found a medium overall level of effectiveness (effect size = 0.637), but highlighted a difference of effectiveness depending on how AR/VR is incorporated into a curriculum. As one example, they report a positive contribution of teaching non-observable subjects (e.g., 3D representations of molecular structures), reducing the rate of error with repetition, and providing innovative, realistic, collaborative, and interactive environments. Similarly, Garzon et al. (2019) in a meta-analysis of AR training interventions found that the effect size varied across the fields of study (e.g., relatively high effect size for training in engineering, manufacturing, and

[1] Reference uses AV as a similar term to VR, which is the consistent term in this report.

construction and a lower effect size when learning about information and communication technologies) and education level (e.g., relatively higher effect size for bachelors-level education versus lower for secondary-level education). As these few examples illustrate, an understanding of AR/VR effectiveness depends on identifying many dimensions that affect their use for training. However, AR/VR reviews to date result in at least as many questions as answers: What is the influence of specific technology characteristics or learning tasks? What are appropriate measures of effectiveness? How does learning acquisition and its persistence vary by learning modality (e.g., visual, cognitive, auditory, motor, sensorimotor coordination)? AR/VR is an exceptionally diverse research field that demands better organization than existing reviews provide.

Another challenge to understanding AR/VR training effectiveness is that the scientists who do the research come from many specializations. They include computer scientists and engineers who focus on technology, educators who focus on instructional methods, and domain specialists (e.g., medical, military, and construction) who focus on implementations in their disciplines. The result is uneven descriptions and findings from one study to another due to the variety of perspectives and their individual, focused goals. We found in an initial review of the literature that some studies lacked details such as descriptions of users' task experience (26% of studies), AR/VR technology experience (28%), and performance measures (29%). We also found inconsistent use of the terms AR and VR themselves and the labeling of technology characteristics, identification of a system without specifying features used, and limited descriptions of tasks trained. To summarize, both the available types of reviews—narrative and meta-analyses—and varied details provided by specific studies, prompt the need for an organizational schema to better capture what is known about the effectiveness of AR/VR.

Rationale for Current Work

There is a considerable amount of knowledge to be gained through conducting effectiveness studies on the use of AR and VR technology for training. However, our examination of AR/VR studies shows that the factors used to determine effectiveness still are being identified and need better categorization. The field requires a more detailed and nuanced organization to better facilitate lessons learned, best practices, and identify gaps.

An important step in understanding a developing area is to properly organize what is known and where there may be knowledge gaps. Part of organizing the field is to develop a taxonomy—a classification system in which components are organized into groups or types based on particular characteristics. The AR/VR framework emerging from our analyses includes: a) the type of technology, b) the skill/task to be trained, c) the trainee, d) the training use or integration into a situation/course, and e) the studies outcome. The development of this framework described next supports a systematic method and format allowing designers and other users to access and understand information they need.

AR/VR FRAMEWORK METHODS AND PROCEDURES

Development of the AR/VR effectiveness framework consisted of three parts: 1) analyzing prior research efforts on published training effectiveness studies, 2) conducting an initial literature assessment that spanned diverse fields and domains of research, and 3) gathering input from AR/VR stakeholders and practitioners. Prior research provided a starting point for the framework by identifying dimensions of AR/VR study characteristics in areas such as maintenance, operations planning, combat, and observation and control (Fletcher et al., 2017). Results of that work suggested that training effectiveness is influenced by task and performance domains (e.g., fast vs. slow movement). A current literature assessment consisted of casting a wide net and sifting through approximately 400 AR/VR effectiveness studies. This assessment revealed additional characteristics of interest as a foundation for the AR/VR framework. Discussions with AR/VR stakeholders and professionals confirmed the need for a systematic analysis and organization of AR/VR studies.

Components of the Framework

The effectiveness framework features four dimensions of AR/VR use and an outcome represented by specific questions to be answered for each AR/VR study as entries for a knowledge base (discussed in the next section). The dimensions and outcome encompass important study information that provides a comprehensive picture of the AR/VR landscape. The framework with brief definitions and associated questions includes:

1. Technology – AR/VR system and components being assessed: What is the device or system being used?
2. Skill/Task – features of the task/job the user learns: What are the tasks/skills being trained or performed with the AR/VR system?
3. Training Use – features of task/job integration: How is the technology used or integrated into a situation/course? What is the training method?
4. Users – description of the trainee's characteristics, the users of the technology: Who are the performers/trainees? What is the prior skill level of user for the task and with the technology?
5. Outcome –assessment of effectiveness: How effective was the AR/VR in supporting performance? What metrics were used to measure effectiveness?

Using the Framework

The AR/VR effectiveness framework aims to compile information from empirical research on training effectiveness. In order to apply the framework to a particular study, two processes take place: 1) answering the framework questions with "specific" and "general" information, and 2) coding each category of the framework with appropriate tags.

Answering Framework Questions

The answers to each of the two framework questions is done in narrative form. The "specific" summarizes the instantiation of the study closely adhering to what in particular was done, how it was done, how it was measured, and its results. The "general" is a description of the characteristics and general implications of the study. For example, the Microsoft HoloLens is a specific technology, while the general description may include the device's characteristics (e.g., see-through head-mounted display (HMD) with head tracking sensors), and how they might be employed in the application. The expectation is that the framework information will provide both an understanding of the study's specific outcomes, and in general how it contributes to a robust and evidence-driven understanding of AR/VR study characteristics and application.

Coding Framework Dimensions

In addition to the narrative descriptions in the framework, each framework dimension of a study is tagged with specific codes. For example, if the technology used in a study was a visual display on a computer monitor or screen the tag = tech_visual_computer, and for an occluded HMD device the tag = tech_visual_HMD_occluded. The tags label common characteristics of studies in the knowledge base and can be used during a query to find studies with the appropriate details. The use of the tags also allows a user to explore gaps (i.e., what tags aren't well populated in the knowledge base) or a specific set of characteristics (e.g., intersection of tasks that use occluded HMD technology and have fast motor movements). The minimum is one tag per category of the framework and applying enough to represent the study.

Building the AR/VR Knowledge Base

Using the framework, AR/VR empirical work is assessed and used to populate a Microsoft Access knowledge base.[2] The knowledge base can be queried by targeting specific descriptions (e.g. MS HoloLens; landing an F-14 on a carrier) or more general characteristics (e.g., see-through head-mounted display; eye-hand coordination). The ultimate goal of the knowledge base is to help tease apart the evidence of AR/VR effectiveness in a more nuanced and systematic way—something that is not done routinely.

The approach of inserting studies into a knowledge base supports many different forms of information, fields, and topics of study. This knowledge base approach helps to identify clusters of studies for users in three broad fields: technology/system developer, training/instructional designer, and instructors who are domain specialists (e.g., flight, maintenance, or construction).[3] Both the kinds of information and its users are essential input for how to organize and query the knowledge base.

---

[2] Microsoft Access is a relational database management system that is part of the Microsoft Office suite included in the Professional and higher editions or sold separately. The content of an Access database can be easily repurposed with a variety of data storage and analytic tools.

[3] No one of these fields is unique, of course, and may be found combined with one another in any specific study.

The result is a multifaceted schema for organizing and coding knowledge base contents. The knowledge base has a standard set of questions and narratives to characterize information, tags to characterize the contents of those narratives, and a search capability to explore varying levels and types of information across studies. The assignment of tags to different kinds of information in studies helps users find what may satisfy their interests when assessing AR/VR training effectiveness. This approach allows us to be more systematic in our analysis of the literature and can capture more of the landscape than a simple keyword search. Particular combinations of characteristics (tags) or of phrases can identify clusters that a keyword search would not (e.g., empirical evidence on hand-held devices and slow motor movement tasks). Searches across studies with similar characteristics can reveal evidence pointing in the same or different direction. Thus, knowledge base users can select the best ways to describe and find what they need.

Identifying and Recording Information in the Knowledge Base

The way to codify the multifaceted schema for organizing and saving information is to develop a template for data entry that encompasses the framework's components. The template describes the information expected in each narrative section, provides space to enter tags for the five components, and includes a place for explanatory notes about the entries. All of the entries and revisions are done with Microsoft Word for ease of entry and editing. The process for evaluating empirical studies includes an initial expert-level review followed by a second expert reviewer to validate the information. The compiled Microsoft Word template is then copied into an Access database that supports search and retrieval functions.

Querying the Knowledge base

The final step for developing the knowledge base is to produce guidelines for users who have varying perspectives and interests, to assist them with compiling the appropriate output from a query. This step includes determining different ways to organize information and providing instructions with examples. We expect differences in what kinds of information will meet the needs of one kind of user versus another such as an instructional designer or a system developer. For example, an instructional designer may initially be most interested in the effectiveness of an AR/VR method proposed by a study, whereas a system developer may want very detailed descriptions of the technology features and capabilities. To date, we have identified a few potential users including system developers, program managers, instructional system designers, trainers/instructors, education and training researchers, and training support personnel.

RESULTS

The result of this effort is an expandable knowledge base: a query-enabled repository of training effectiveness studies. The knowledge base allows users to search for studies based on shared characteristics (e.g., systems that include an occluded HMD and automated feedback as part of the training method). Such searches are satisfied based on the structural and searchable features of the knowledge base—the key to meeting users' needs. At the time of writing this paper, the knowledge base holds summaries of 64 effectiveness studies. That repository represents a cross-section of publications mostly from the past five years with the near-term plan for it to expand over time.

AR/VR Technologies

The knowledge base supports a broad range of queries and perspectives. For example, a user may ask what AR/VR hardware is addressed in the knowledge base? Table 1 shows the frequency of the 10 most common AR/VR technologies (i.e., hardware) in the knowledge base studies. The different kinds are an indication of the variety encompassed by AR or VR; there are additional technologies in the knowledge base, but we limited the table to the top 10 for brevity.

The most frequent technology is an occluded HMD, which appeared in half of the articles. Other visual display technologies include computer monitors, see-through HMDs, projections on a surface, and handheld tablets. These visual displays span the continuum of mixed-reality display types as described by Milgram and Kishino (1994). The knowledge base also includes other technologies such as systems that track a user's hands or head, headphones for providing audio information to the user, and systems for tracking equipment that a user manipulates or items in the real world that may have markers (e.g., QR codes) for identification.

Table 1. The 10 Most Frequent Hardware Technologies Assessed in the Knowledge Base

Technology	Count of Studies
Occluded HMD	32
Tracking user's hands (optical hand tracking or device in/on hands)	31
Tracking users head	24
Computer monitor used for visual display	18
See-through HMD	14
Headphones to provide audio to user	15
Tracking system for equipment or mechanical device	12
Tracking system for applied markers to registered items	8
Projector provides visual information to user	7
Handheld tablet or phone provides visual display of information	7

If a knowledge base user were interested in a particular technology, a query that specified that technology application would produce all of the studies in the knowledge base that included effectiveness results of how that technology was used. For example, if a training developer were interested in how an occluded HMD could be used, a search would show the summaries of all 32 studies in the knowledge base. By going through the summaries, the training developer would learn about the different ways that such a device has been used effectively, such as: improving performance of novice robot teleoperators (Brizzi et al., 2018); increasing safety for crane operators (Dhalmahapatra et al., 2021); and increasing students' ability to recognize and correct problems while parachuting (Liang et al. 2020). This query would also provide some cases where the VR training was less effective than the traditional training method (e.g., Winther et al., 2020; Oberhauser et al., 2018), and reading those summaries may provide the knowledge base user with ideas on what problems to avoid. Additionally, there are studies that assessed the subcomponents of a system to show which system configurations worked relatively better or worse, like Westerfield et al. (2013) demonstrating how the addition of an intelligent feedback system improves the training of assembly tasks, and Monteiro (2020) showing how different sensory cues influence performance in firefighting training. Knowledge base searches of any of the framework dimensions can provide a training developer with ideas about how the searched-for category can be used and under what conditions it might be effective.

While the above examples illustrate how a user query of the knowledge base can find existing empirical evidence, it can also help identify where gaps exist. Identifying knowledge gaps can help users of the knowledge base prioritize the research and development needs in the field. For example, currently there are only single studies in the knowledge base that assess systems that track the user's eye/gaze or use haptic gloves to provide feedback; and only two studies assess technologies that autonomously identify the surrounding environment. Such findings can alert technology developers to the need for new research while trainers can identify what is not supported by rich empirical evidence.

Study Outcomes

A knowledge base user's approach to outcomes and its measurement can be from different perspectives. For example, trainers may assess performance on a physical task, testing cognitive understanding and knowledge gain, or rating how well a system can be used or the level of work required to use a device. All of those scores provide a quantifiable means for assessing outcomes. There are also qualitative methods for assessing outcomes, like the opinions of experts or system users to describe how successful a system might be for accomplishing a particular task. The knowledge base includes codes that characterize the type of information (e.g., objective empirical measure of performance, subjective rating of performance or expert opinion) in each study to describe its method for assessing outcomes.

Assessing how well a trainee physically performs the actual training task is a clear way to determine training effectiveness. Figure 1 shows the number of studies in the knowledge base, grouped by how they measure the training task and whether the finding indicated if the AR/VR system was more or less effective than traditional training, if results were mixed, or if the assessment compared system components to one another. In the knowledge base, Results show that 38 of 64 studies include objective measures of how well the users physically performed the training task (e.g., time, error rate) as indicated by the blue bars. Fourteen studies assessed physical performance by subjective ratings (e.g., Likert scale) of how well a trainee performed as indicated by the orange bars, and 20 studies assessed trainees through the use of cognitive measures as shown by the gray bars. Of the 38 studies that objectively measured performance, 14 demonstrated that training with the AR or VR system improved performance more than a traditional training method, 8 studies indicated mixed results (i.e., some aspects of performance were better, worse, or not significantly different), and only 2 studies showed that AR/VR training was less effective than the traditional methods. When physical performance was rated, the results were relatively similar with the highest frequency being studies that show AR/VR to be more effective than traditional methods, followed by mixed results; no studies indicated the system was less effective. When cognitive measures were used, the most frequent finding was mixed results, followed by AR/VR being more effective, and only two studies where AR/VR was less effective. Overall, the knowledge base reveals that AR/VR is usually more effective than traditional methods or, at least, that the results are mixed, with only a few studies indicating that AR/VR is less effective than traditional methods. In general, these are findings similar to what might be in a meta-analysis, comparing AR/VR to a traditional method for training.

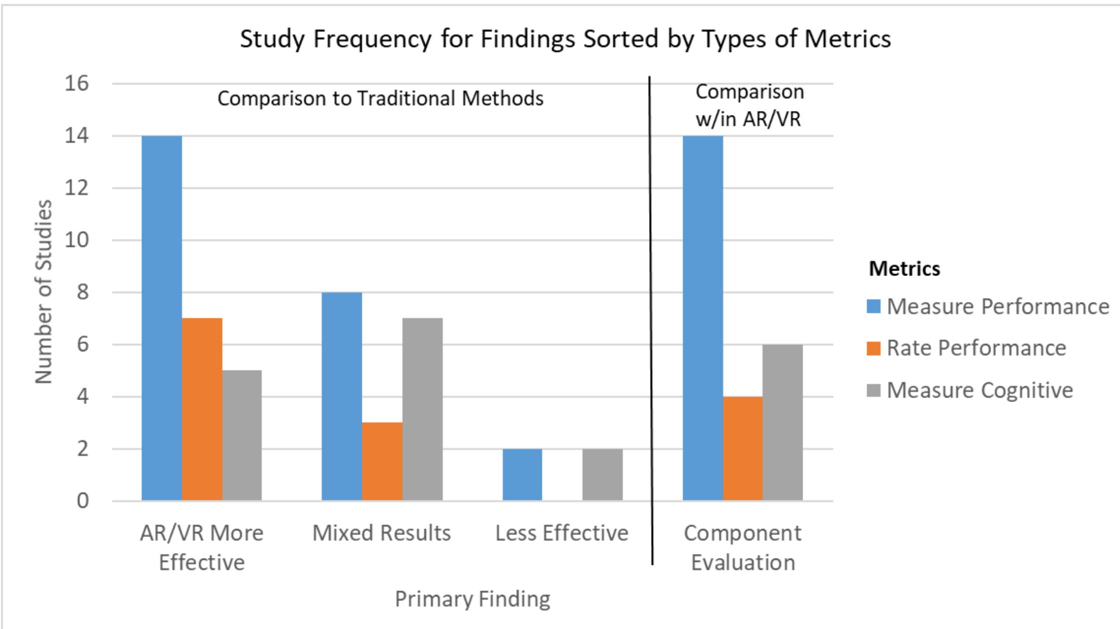

Figure 1. The number of studies that assess a trainee's ability on the training task, grouped by the study findings demonstrating a system was more or less effective, if the results were mixed, or if the study compared components of the system.

Shown in the right section of Figure 1, there are 24 studies that assessed components within an AR/VR system (i.e., no comparisons to traditional or a non-AR/VR method). These studies attempt to identify the relative benefits of particular configurations of the system. A mix of metrics were used across these assessments, with measuring physical performance being the most common, followed by measuring cognitive outcomes, and finally rating performance measures. This is information that would not be included in a meta-analysis in that there is no comparison to traditional methods. However, this is important information that is captured in the knowledge base so that users can learn what components of AR/VR may be more or less beneficial in particular training contexts.

User/Trainee's Experience

Understanding the initial state of the user (i.e., trainee) prior to the training is important because prior experience with virtual environments and related technology has been shown to influence subsequent performance (Smith & Du'Mont 2009; Orvis et al., 2005). In the current knowledge base, only 38 of 64 studies that describe the user's experience level with the technology or the training task. Of the 38 studies that provided information to assess trainee's prior experience level, Figure 2 shows the number of studies and their type of findings (i.e., comparison to traditional method or comparison within AR/VR system) as grouped by the user experience level (i.e., a novice or experienced with either the training task or with AR/VR technology). The majority of these studies are on trainees who are novices with the technology, shown by the blue (also a task novice) and the orange (prior task experience) bars. Being new to the technology could influence how well a trainee does to learn skills/tasks with AR/VR.

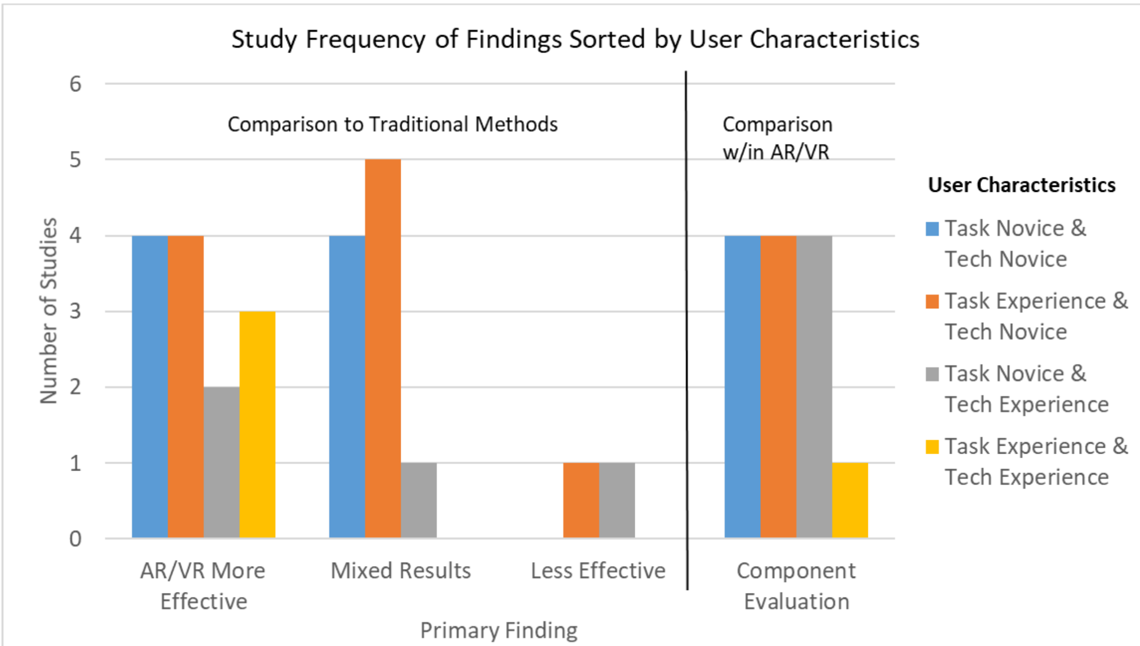

Figure 2. The number of studies with trainees grouped by their level of prior task and technology experience.

Tasks/Skills

The knowledge base includes training effectiveness studies where users perform different types of tasks or learn different types of skills. One way to characterize the performance of users is by describing the overall field or training domain; another way is by breaking down the training to fundamental types of actions. When characterizing each study as a singular field or training domain, the most common are medical procedures (12 studies), repairing/assembling equipment (11), construction and building maintenance tasks (5), and robotics, military actions, and pathfinding with 3 studies each. For training developers who focus on a particular field, they can find in the knowledge base how much work is being done.

Knowledge base queries also support deeper dives into training effectiveness topics such as the fundamental types of actions/activities that cut across training domains. These actions/activities are multidimensional including physical (manipulating objects) and cognitive control functions (decision-making, acquiring knowledge) among many others. Categorizing the studies by these fundamental actions is helpful because there may be lessons learned across domains (e.g., medical and construction) that can be more easily found because of the similarities in the fundamental actions. For example, training a person to visually discriminate a disease in medicine may share some characteristics with training a person to visually discriminate a construction problem. The performance codes allow a search to reveal the many kinds of crossovers. The most frequent include eye-hand coordination (24 studies), visual discrimination of objects (24), acquiring knowledge (21), fine (detailed) manipulation of objects (15), and decision making (10). Characterizing such training effectiveness actions/activities may allow someone in a field that has not used AR/VR for training to see how their interests link to another field.

Intersection of Framework Components

Similar to cutting across training domains, an important characteristic of the knowledge base is that users can find the intersection of framework dimensions as part of queries. For example, the trainer can not only see the range of technologies included in AR/VR training effectiveness studies, but also can drill down deeper into how a particular technology might be used with a specific training method as part of a training system.

For a sample query, "How can occluded head mounted visual displays (i.e., a particular technology) be incorporated into training with the inclusion of automated feedback (i.e., a particular training method)?" the knowledge base returns eight studies. The findings provide a combination of technology and feedback mechanisms that trainers/training developers can probe to identify relevant findings for their use case. Two studies that compared the VR systems to traditional training had different results. One found that an immersive VR system outperformed conventional training on two behavioral transfer tests (Makransky et al., 2019); the other demonstrated that a developmental version of VR training was effective but not as effective as the well-developed traditional training system (Winther et al., 2020). Four other studies compared different features of VR training to one another, and those findings included: the addition of intelligent tutoring capabilities improve training outcomes over VR training without intelligent feedback (Westerfield et al., 2013); the types of feedback from movement/locomotion in virtual training influenced system usability and performance (Corelli et al., 2020); and matching types of sensory stimulation for feedback influences training effectiveness (Batmaz & Stuerzlinger 2021; Monterio et al., 2020). Lastly, there were two studies describing the benefits of VR training with automated feedback, such as decreasing training costs and enhancing training safety (Kwegyir-Afful & Kantola 2021; Song et al., 2021).

By combining framework components and searching for nuances, trainers can probe to a level-of-detail up to and including individual studies. They can identify relevant information for their specific needs or determine that recent AR/VR studies may not be relevant. They can query the knowledge base by combining its many different dimensions in cross-walks of technology type, tasks trained, training methods used, participants, and effectiveness. The knowledge base allows its users to personalize their AR/VR searches.

Conclusion

The state of training effectiveness studies on AR/VR technology is currently fragmented and disorganized. Some of the reasons for this disorganization is the inconsistent use of terms and studies being done by researchers in disparate fields (e.g., technologist, education/training researchers, and domain specialists). The development of a framework helps to organize the information that may influence or impact the effectiveness of a training system. Then, by applying that framework to summarize training effectiveness studies, the content for a knowledge base is created. This knowledge base with organized content allows users to filter for one or more information categories that address their specific queries.

ACKNOWLEDGEMENTS

This effort was funded by the Air Force Research Laboratory, Airman Systems Directorate, Warfighter Interaction and Readiness Division (AFRL/RHW) in support of the Commander Air Combat Command's (COMACC) migration to Proficiency-Based Training.

REFERENCES

Batmaz, A. U., & Stuerzlinger, W. (2021, March). The Effect of Pitch in Auditory Error Feedback for Fitts' Tasks in Virtual Reality Training Systems. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR*), 2021, pp. 85-94. doi: 10.1109/VR50410.2021.00029

Batdi, V., & Talan, T. (2019). Augmented reality applications: A Meta-analysis and thematic analysis. *Turkish Journal of Education*, 8(4), 276-297.Brizzi, F., Peppoloni, L., Graziano, A., Di Stefano, Erika; Avizzano, C. A., & Ruffaldi, E. (2018). Effects of Augmented Reality on the Performance of Teleoperated Industrial Assembly Tasks in a Robotic Embodiment. In *IEEE Transactions on Human-Machine Systems*, 48(2), (2018): 197-206. DOI:10.1109/THMS.2017.2782490

Corelli, F., Battegazzorre, E., Strada, F., Bottino, A., Cimellaro, G. P. (2020). Assessing the usability of different virtual reality systems for firefighter training. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2, pp. 146-153.

Dhalmahapatra, K., Maiti, J., Krishna, O. B. (2021). Assessment of virtual reality based safety training simulator for electric overhead crane operations. *Safety Science*, 139, art. no. 105241.

Fletcher, J. D., Belanich, J., Moses, F., Fehr, A., & Moss, J. (2017). *Effectiveness of augmented reality & augmented virtuality*. In MODSIM (modeling & simulation of systems and applications) World Conference. https://modsimworld.org/papers/2017/Effectiveness_of_AR_and_VR.pdf.

Garzón, J., Baldiris, S., Gutiérrez, J., & Pavón, J. (2020). How do pedagogical approaches affect the impact of augmented reality on education? A meta-analysis and research synthesis. *Educational Research Review*, 31, 100334.

Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of Augmented Reality on students' learning gains. *Educational Research Review*, 27, 244-260.

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors*, 63(4), 706-726.

Kwegyir-Afful, E., Kantola, J. (2021). Simulation-Based Safety Training for Plant Maintenance in Virtual Reality. *Advances in Intelligent Systems and Computing*, 1206 AISC, pp. 167-173.

Liang, C. Y., Lascink, R., & Klyde, D. H. (2020). Operational Evaluation of a Virtual Reality Parachute Simulator. In *AIAA Scitech 2020 Forum* (p. 0669). https://doi.org/10.2514/6.2020-0669.

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning*, 35(6), 691–707.

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

Monteiro P., Melo, M., Valente, A. Vasconselos-Raposo, J. & Bessa, M. (2020). Delivering Critical Stimuli for Decision Making in VR Training: Evaluation Study of a Firefighter Training Scenario. In *IEEE Transactions on Human-Machine Systems*, 51(2), (2018): 65-74. DOI: 10.1109/THMS.2020.3030746.

Oberhauser, M., Dreyer, D., Braunstingl, R., & Koglbauer, I. (2018). What's real about virtual reality flight simulation?: Comparing the fidelity of a virtual reality with a conventional flight simulation environment. *Aviation Psychology and Applied Human Factors*, 8 (1): 22-34. DOI:10.1027/2192-0923/a000134.

Orvis, K. A., Orvis, K. L., Belanich, J., & Mullin, L. N. (2005). *The influence of trainee gaming experience and computer self-efficacy on learner outcomes of videogame-based learning environments*. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Smith, S. P., & Du'Mont, S. (2009, March). Measuring the effect of gaming experience on virtual environment navigation tasks. In *2009 IEEE Symposium on 3D User Interfaces* (pp. 3-10). doi: 10.1109/3DUI.2009.4811198

Song, H., Kim, T., Kim, J., Ahn, D., & Kang, Y. (2021). Effectiveness of VR crane training with head-mounted display: Double mediation of presence and perceived usefulness. *Automation in Construction*, 122, 103506.

Westerfield, G., Mitrovic, A., & Billinghurst, M. (2013, July). Intelligent augmented reality training for assembly tasks. In *International Conference on Artificial Intelligence in Education* (pp. 542-551). Springer, Berlin, Heidelberg.

Winther, F., Ravindran, L., Svendsen, K. P., & Feuchtner, T. (2020, April). Design and evaluation of a VR training simulation for pump maintenance. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-8). https://doi.org/10.1145/3334480.3375213