Examining Enhanced Learning Diagnostics in Virtual Reality Flight Trainers

Gregory McGowin[1], Zerong Xi[1], Olivia B. Newton[1], Gita Sukthankar[1,]
Stephen M. Fiore[1], and Kevin Oden[2]
[1] University of Central Florida US
[2] Lockheed Martin, Orlando FL US

As the complexity of aircraft cockpit operations increases, training effectiveness must be improved, and learning accelerated. Virtual reality (VR) training is increasingly offered as a method for improving training efficacy given its ability to provide a rich sensory experience during learning. This paper describes a study examining how training efficacy can be improved by improving learning diagnostics. We study how varying forms of knowledge assessment are related to different types of task knowledge and task performance in a VR flight simulator. The data suggest that participants who demonstrated higher training comprehension, measured via diagnostic test questions, on conceptual (and to a lesser effect) declarative knowledge, also demonstrated superior knowledge transfer in the VR flight simulator. Findings are discussed in the context of improving cognitively diagnostic assessments that are better able to predict task performance and inform individually tailored training remediation.

INTRODUCTION

Broadly construed learning involves not simply the acquisition of new knowledge, but also the process through which individuals construct knowledge via the active integration of concepts. Definitions of learning vary, often depending on discipline or research context. We adopt the perspective coming out of cognitive psychology that defines learning as "the relatively permanent changes in behavior or knowledge that support long-term retention and transfer" (Soderstrom & Bjork, 2015, p. 176). While learning can take place in a multitude of settings, today, learning often takes place in a multimedia context. This form of multimodal learning results from an interaction between the student and some multimedia interface, such as computer-based training or smartphones (e.g., Castro-Alonso, 2019). Here, the use of different media formats is postulated to enhance the learning experience.

Prior research has sought to classify media formats into three distinct levels of interactivity: low, medium, and high. Low interactivity refers to a simple dialogue between the student and multimedia, medium interactivity provides the learner with some degree of agency (e.g., controlling the pace of a video via pause/rewind buttons), and high interactivity allows the learner to directly alter media formats (Castro-Alonso & Fiorella, 2019; Moreno & Mayer, 2007). While much research has explored the variety of multimedia methods available for learning (e.g., virtual learning environments, webinars, recorded videos), virtual reality simulators are being implemented in learning environments at a pace that exceeds our ability to understand how to optimize their implementation. Although virtual reality has a long history of research in application, theory and methods from the learning and cognitive sciences have yet to be fully studied with virtual reality. As such, interdisciplinary research is needed to more fully examine the ways in which virtual reality can augment traditional training. In this paper we first describe some of the basics of virtual reality technology. We then summarize findings from learning and training research that can be adapted to virtual reality so as to better understand the relationship between learning and transfer. Finally, we describe a study where these methods were integrated so as to study learning and training using a blend of traditional media with virtual reality simulation assessments.

Virtual reality (VR) has had many different definitions across different domains (e.g., LaVelle, 2019; Briggs, 1996; Kizilkaya et. al, 2019), however it is generally understood as the use of 3D graphic systems that digitally simulate or replicate an environment (Makransky & Lilleholt, 2018) to provide the effect of immersion in an interactive virtual environment (Pan et al., 2006). These computer-simulated environments induce "targeted behavior in an organism by using artificial sensory stimulation, while [optimally] the organism has little or no awareness of the interference" (LaValle, 2019 p. 1).

Virtual reality studies have shown several positive benefits for learning and training compared to less interactive media. Markowitz et. al (2018) demonstrated that VR can improve knowledge gains as well as improve interest in STEM topics (i.e., climate science), while other studies have demonstrated positive changes in self-efficacy, increased knowledge acquisition, and transfer (Meyer et. al, 2019; Makransky et. al, 2019), along with improved recall (Krokos et. al, 2019). Virtual reality simulations are thought to facilitate the learning process by placing the trainee in a performance environment where they engage in activity that provides experiential practice. Before the advent of simulations, the only way to train what is called a "learning-to-do skill" (Aldrich, 2009) was to go out and "do" the skill. Acquisition of knowledge was accomplished by attempting the skill over and over again. With the advent of simulations, these "learning-to-do skills" can now be replicated not just in physical environments, but in virtual reality environments as well.

Simulations also provide a broader swath of learning diagnostics given that numerous performance indicators are collected. Because of this, the training context allows instructors to more precisely target where learning deficiencies

might be occurring. Additionally, VR provides a safe and cost-effective means to engage in the repetitive practice needed to accelerating learning complex procedural skills that require integration of cognitive and motor knowledge. The affordability of virtual reality has increased the use of these immersive environments with the hope that they can enrich learning experiences and increase knowledge transfer (e.g., Makransky, Borre-Gude, & Mayer, 2019; Makransky & Petersen, 2019; Meyer, Omdahl, & Makransky, 2019; Parong & Mayer, 2018). Moreover, simulated scenarios, VR-based or otherwise, are easily repeated without consequence. Compared to, for example, real-world flight, which is cost and time intensive, these work well for tasks that require repetition (e.g., aircraft takeoff/landing, target practice). Further, they require less reset time, allowing for more efficient simulation training (Aldrich 2009). This means that training time is devoted to the actual task, and not ancillary tasks (e.g., circling the airfield before restaging an attempt to land). Importantly, this allows for easy implementation of multiple tests. This can accelerate learning because, over and above training and repetitive practice, testing is known to influence individual learning (McDaniel, Roediger, & McDermott, 2007).

In combination, virtual reality simulations provide a powerful means for examining learning and training. When considering VR simulators for testing, they may provide an additional diagnostic capable of better predicting training efficacy. For example, previous research shows that different types of knowledge acquisition assessments predict transfer performance on complex decision-making tasks (Fiore et al., 2002; Song et al., 2018). Recently, Song and colleagues (2019) found that training comprehension was predictive of complex transfer-type tasks assessing decision making under uncertainty. More succinctly, the learner's ability to acquire and integrate concepts from training was predictive of performance in transfer tasks requiring application of this knowledge. These methods are foundational to training efficacy because well-designed training assessments, capturing precisely how knowledge is acquired, can improve overall training effectiveness. Specifically, through the use of such approaches, trainers may be better able to diagnose learning progress by pinpointing the particular elements of knowledge that have been acquired (see Cuevas et al., 2004). From this, then, training precisely tailored for individual learners is possible.

In this study, we extend prior work by using virtual reality simulations as our transfer of learning context. We examine the processes by which individuals acquire and integrate knowledge with multimedia learning and assess application in an immersive virtual environment. Specifically, to explore the effects of acquiring varied knowledge types on simulator performance, we developed training for a VR flight simulator and assessed three types of knowledge: recognition, declarative, and conceptual. We hypothesized that the amount and type of knowledge acquired would be differentially predictive of task performance in the VR flight simulator.

METHODS

Study Design

For this experiment, we used a within-subjects research design, specifically varying the difficulty of the flight scenarios used in the VR environment. All participants went through a training tutorial on the basics of flight and their knowledge acquisition was assessed. They all also engaged with the same set of practice and test flight scenarios, but the presentation order for the scenarios was randomized using Latin Square counterbalancing. The VR flight scenarios contained three levels of relative difficulty (i.e., easy, medium, and hard). Other data were collected (e.g., eye tracking, a cognitive battery, workload) but, due to space limitations, we report only on the results relating to tests of knowledge acquisition and VR simulator performance.

Participants

The sample consisted of 23 undergraduate students, 15 male (65%) and 8 female, ages 18-29 (M = 19.5, SD = 2.4) recruited from a southeastern U.S. university. Participants were considered naive pilots, with the majority (87%) having little to no familiarity with flight simulators. Most participants (78%) rated their video game expertise as fair or less and played video games an average of 7.5 hours (SD = 8.4) a week. Students were compensated with course credit for participating in the study.

Procedure

Participants received a self-paced tutorial that included both text and images on the basics of flight. The tutorial introduced the participants to the aircraft, controls (throttle and joystick), basic flight maneuvers (i.e., ascending and descending, increasing and decreasing speed, turns, and straight-and-level flight), and to the aircraft's flight instruments (i.e., airspeed indicator, attitude indicator, altimeter, vertical speed indicator, heading indicator, turn coordinator, and stall-warning indicator). All participants received the same training administered via an online survey tool and completed it in approximately 14 minutes (M = 13.85 min., SD = 6.6 min.).

Knowledge Acquisition Assessment

The knowledge acquisition assessment consisted of three sets of test items: 1) recognition, 2) declarative, and 3) conceptual. *Recognition* knowledge is developed via repeated exposures to pictorial/visual task relevant stimuli. The assessment consisted of multiple-choice questions in which participants were shown an aviation related image and needed to correctly match it to the name of the airplane part or instrument. *Declarative* knowledge represents learners' mastery of basic factual concepts associated with the training. It was assessed with multiple-choice, true/false, as well as text entry questions. *Conceptual* knowledge develops through the integration of training relevant concepts. Here, training content is synthesized into meaningful knowledge structures that associate different concepts. This represents a more complex form of knowledge. It was evaluated

with multiple- choice questions requiring participants to accurately integrate concepts and identify which set are required for a correct answer. Examples of these questions are presented in Table 1.

Table 1. Example Knowledge Assessment Questions.

Recognition	This instrument is the _____. [image of flight instrument such as altimeter]
Declarative	What altitude is being shown on this altimeter?
Conceptual	Using the current state of the attitude indicator, what actions do you need to perform to return the aircraft to straight and level flight (Figure 1). a) Pitch: No change, Bank: Roll left; b) Pitch: No change, Bank: Roll right; c) Pitch: Nose tilting up, Bank: No change; d) Pitch Nose tilting down, Bank: No change

Figure 1. Image of an attitude indicator used in conceptual assessment.

VR Flight Simulator

Participants used a VR capable flight simulator (Prepar3D®) to perform nine randomized practice flight scenarios and nine randomized test flight scenarios. Table 2 illustrates scenario difficulty, time limits, and task requirements. Both the VR practice and test flight scenarios took approximately 10 minutes each to complete.

During each scenario, the built-in programs of Prepar3D collected aircraft status data for each participant. This aircraft status data reported the aircraft's coordinates, altitude, orientation (i.e., pitch, yaw, roll), and velocity, as well as controller input (e.g., percentage of aileron, elevator, and throttle). This data was used to calculate success/failure based on a criterion that the aircraft maintain a target status for five or more continuous seconds (e.g., maintained a specified altitude given in the instructions for 5+ seconds). The test scenarios were considered the transfer task as they required integration and application of knowledge acquired from the tutorial but executed in the virtual reality simulator. Performance in these test scenarios was the primary dependent variable.

Table 2. Flight simulator scenarios.

Difficulty	**Time** (sec)	**Direction Change** (degrees)	**Reduce Airspeed** (mph)	**Altitude change** (ft)
Easy	60	90	0	0
	60	0	50	0
	60	0	0	-500
Medium	80	-90	0	0
	80	0	50	0
	80	0	0	-1000
Hard	100	90	0	2000
	100	-90	0	0
	100	0	50	-1000

Apparatus

The tutorial and assessment were presented through Qualtrics, a presentation and survey tool on a 22-inch monitor with 1920 x 1080 resolution at 60 Hz. The 3-D flight simulator used to present the simulation and capture the simulator data was Prepar3D, developed by Lockheed Martin Co. Participants interacted with the flight simulator via the HTC Vive Pro, a VR head mounted display (HMD). This had a built-in Tobii eye tracker, with a resolution of 1440x1600 per eye, a refresh rate of 90Hz, and a field of view (FOV) of 110 degrees. Participants flew the plane using the Logitech x56 HOTAS (joystick and throttle). The VR flight simulator was run on a high-performance consumer grade PC (CPU = i7-9700k, GPU = RTX 2080ti).

RESULTS

Descriptive Statistics

Participant accuracy in the knowledge tests was relatively high. On the Recognition knowledge tests, mean accuracy was .84 (*SD* = .19). For the Declarative knowledge tests, mean accuracy was .86 (*SD* = .18). Last on the Conceptual knowledge tests, mean accuracy was .79 (*SD* = .20). For the VR flight scenario tests, across levels of difficulty, participants achieved a mean success rate of .40 (*SD* = .18).

Correlations

We first wanted to assess the relationship between knowledge acquisition and knowledge application. Knowledge acquisition is determined by accuracy on the differing knowledge tests and knowledge application is assessed via performance on the VR test flight scenarios. Correlations with these measures were run (see Table 3) across these measures. Looking across these items, we see that knowledge acquisition between Declarative and Recognition tests was significantly correlated. But only Declarative knowledge was significantly correlated with the Conceptual knowledge tests. Further, both Declarative and

Conceptual knowledge tests were significantly correlated with performance on the test flight simulator scenarios.

Table 3. Pearson Correlation Coefficients for VR Test Scenario Performance and Knowledge Acquisition Assessments.

	Simulator Perf	*Rec.*	*Dec.*	*Conc.*
Simulation Performance	1			.
Recognition	0.152	1		
Declarative	.428*	.695**	1	
Conceptual	.513*	0.402	.668**	1

*$p \leq 0.05$ level, **$p \leq 0.01$ level (2-tailed).

Linear Regression

We next wanted to directly test the relationship between the varied knowledge tests and performance on the flight simulator tests. Again, the goal here is to determine what forms of knowledge are differentially diagnostic of actual operational performance; that is, how can knowledge acquisition be used to predict knowledge application on transfer tasks assessed with VR. Simple linear regressions were calculated to predict flight simulator performance based on accuracy on the knowledge acquisition tests. For the conceptual knowledge assessment, a significant regression equation was found ($F(1,21) = 7.5$, $p < .01$, $R^2 = 0.26$). Participants predicted flight simulator performance is equal to 0.58 + .55 (conceptual) score. For the declarative knowledge assessment, a significant regression equation was found ($F(1,21) = 4.7$, $p < .04$, $R^2 = 0.18$. Participants predicted flight simulator performance is equal to 0.69 + 0.42 (declarative) score. For the recognition knowledge assessment, a non-significant regression equation was found ($F(1,21) = 0.5$, $p < .48$, $R^2 = 0.02$.

DISCUSSION

This study set out to examine how virtual reality flight simulator scenarios could be used to better diagnose learning efficacy. A complex task, flight maneuvers, was used as the learning environment. Operating an aircraft is a complex skill that requires knowledge of instrumentation, control inputs, and flight maneuvers, to name a few. Being able to take these different aspects of aviation and merge them together is what allows novice pilots to be successful. In line with previous research (Cuevas et al., 2002; Fiore et al., 2003), participants who, not only better acquired, but also better integrated concepts from the learning tutorial, also demonstrated superior knowledge transfer in the VR flight simulator. As with prior research, the simplest test type (Recognition knowledge), was not predictive of VR flight scenario performance. However, for the next most difficult (Declarative knowledge), we do find predictive utility. But the most predictive, based upon the R2 change, was the conceptual knowledge test. This provides evidence that conceptual tests are able to tap into a deeper form of knowledge integration, and can be more diagnostic of learning efficacy, in this case, predictive of performance for aircraft operations in a VR setting.

These findings contribute to both basic and applied research in training. First, from the applied perspective, this helps training designers target different forms of knowledge appropriate for differing levels of expertise. Too often, assessments are based solely on being able to correctly recognize an object, or to provide basic factual knowledge; that is, only test recognition and declarative knowledge. As such, they can miss the crucial step of integration of the information. This is of critical importance as complex real- world tasks, such as operating an aircraft, require a person to have integrated multiple concepts together to form meaningful knowledge structures of how to perform that task successfully. With these findings, we provide evidence that such measures can be differentially predictive of more complex transfer forms of knowledge applications as assessed via virtual reality flight simulations. Finally, from a fundamental science perspective, this research helps us understand the foundational forms of knowledge acquisition learners may go through as they move across levels of expertise. This, then, can better elucidate how training research can tailor the learning environment to potentially accelerate the development of expertise for individual trainees (Hoffman et al. 2013).

In sum, this study set out to determine if differing forms of knowledge assessment could predict performance in a complex VR flight simulator. A computer-based training tutorial was developed around the principles of flight and associated knowledge tests were developed (cf. Fiore et al., 2003; Cuevas et al., 2004). Following prior research on cognitively diagnostic assessment (Cuevas et al., 2004; Fiore et al., 2002), these tests tapped the acquisition of relatively rudimentary knowledge (recognition tests), as well as more challenging forms of knowledge (declarative knowledge). Most importantly, a form of knowledge acquisition requiring the integration of concepts was also developed. This represents a higher form of understanding in that, in order to be accurate, participants had to synthesize the relationship across items. Our results show that tests designed to capture more complex forms of knowledge acquisition are better able to predict transfer tasks more similar to operational type performance. As such, this research can be used to better design training that more accurately diagnoses learning deficiencies and, from this, identifies appropriate training remediation.

This study extends the knowledge base for both the learning and cognitive sciences by demonstrating how a blend of traditional media with virtual reality can enhance conventional training. It also provides additional insight into the relationship between learning and transfer in virtual reality.

AKNOWLEDGEMENT

This study was supported by funding from Lockheed Martin Corporation to Gita Sukthankar and Stephen M. Fiore, with Kevin Oden as program manager. The views and opinions contained in this article are the authors' and should not be construed as official or as reflecting the views of the University of Central Florida or Lockheed Martin Corporation.

REFERENCES

Aldrich, C. (2009). *The Complete Guide to Serious Games and Simulations*. Wiley.

Briggs, J. C. (1996). The promise of virtual reality. *The Futurist, 30*(5), 13-18.

Castro-Alonso, J. C., & Fiorella, L. (2019). Interactive science multimedia and visuospatial processing. In *Visuospatial processing for education in health and natural sciences* (pp. 145-173): Springer.

Cuevas, H. M., Fiore, S. M., Bowers, C. A., & Salas, E. (2004). Fostering constructive cognitive and metacognitive activity in computer-based complex task training environments. *Computers in Human Behavior, 20*, 225-241.

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes: Use of diagrams in computer-based training environments. *Instructional Science, 30,* 433 – 464.

Fiore, S. M., Cuevas, H. M., & Oser, R. L. (2003). A picture is worth a thousand connections: The facilitative effects of diagrams on task performance and mental model development. *Computers in Human Behavior, 19*, 185-199.

Fiore, S. M., Cuevas, H. M., Scielzo, S., & Salas, E. (2002). Training individuals for distributed teams: Problem solving assessment for distributed mission research. *Computers in Human Behavior, 18*, 125 - 140.

Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D. H. (2013). *Accelerated Expertise: Training for High Proficiency in a Complex World.* East Sussex, UK: Psychology Press

Krokos, E., Plaisant, C., & Varshney, A. (2019). Virtual memory palaces: immersion aids recall. Virtual Reality, 23(1), 1-15

LaValle, S. (2019). Virtual Reality. Retrieved from http://vr.cs.uiuc.edu/

Markowitz, D. M., Laha, R., Perone, B. P., Pea, R. D., & Bailenson, J. N. (2018). Immersive virtual reality field trips facilitate learning about climate change. Frontiers in Psychology, 9, 2364.

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

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. Educational Technology Research and Development, 66(5), 1141-1164.

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14(2),* 200-206.

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, 103603.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309-326.

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. Journal of Educational Psychology, 110(6), 785.

Pan, Z., Cheok, A. D., Yang, H., Zhu, J., & Shi, J. (2006). Virtual reality and mixed reality for virtual learning environments. Computers & graphics, 30(1), 20-28.

Soderstrom, N.C. & Bjork, R.A. (2015). Learning versus performance: An integrative review. Perspectives on Psychological Science, 10, 176–199.

Song, J., Newton, O. B., Fiore, S. M., Coad, J., Clark, J., Pittman, C., & LaViola, J. J. (2018). Examining the Impact of Training and Feedback on Visualization-Supported Decision Making under Uncertainty. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,, 62*(1), 1449-1453.

Song, J., Newton, O. B., Fiore, S. M., Pittman, C., & LaViola, J. J. (2019). Examining Training Comprehension and External Cognition in Evaluations of Uncertainty Visualizations to Support Decision Making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63(1),* 1654–1658.