

## Quantitative Analysis of Virtual-Reality Training Effectiveness for Cockpit Procedures

**Mark K. Budgeon, Brandon S. Wolf, Margaret T. Merkle, Donna C. Senft**  
mark.bugeon.2@us.af.mil, brandon.wolf.6@us.af.mil, margaret.merkle.1@us.af.mil, donna.senft@us.af.mil

### ABSTRACT

Recent advances in computers and technology have made VR (Virtual Reality) devices with imbedded training applications more commercially available. Additionally, there is potential to realize effective training with minimal human-instructor interaction. Our hypothesis is that a VR training environment, coupled with software designed to effectively guide the subject through a task, would improve the performance outcomes of a USAF pilot measured by a reduction in time to complete a task or reduction in errors. The authors chose checklist execution as the measured task during static flight operations (e.g. the aircraft may be “airborne” but the focus is on checklist execution rather than aircraft control), and investigated whether training provided by the VR device improved pilots’ performance. Twelve subjects (n=12) were selected; six had no VR training (no-VR) and six had VR training (yes-VR). The six no-VR subjects had completed a 3-month training program consisting of lectures, computer-based training, testing, and guided practice sessions at a physical cockpit station (aka “cardboard trainer”). Whereas, the six yes-VR subjects were only afforded 20 hours of training exclusively within the VR device. They had no other academics. After completing training, both the no-VR and yes-VR groups were asked to complete the BEFORE STARTING ENGINES CHECKLIST in a new environment – a full-motion simulator representative of the actual aircraft. The study used measures “time to completion” (time) and “number of errors” (numErr) as indicators of subject pilot performance. The results showed that there was no statistically significant difference in time between the no-VR and yes-VR subjects ( $p=0.8713$ ). However, the number of mistakes made by the yes-VR subjects was significantly less than the no-VR subjects ( $p=0.0402$ ). The authors conclude that a VR device coupled with software designed to focus the subject’s training to the task, is responsible for the reduction in errors.

### ABOUT THE AUTHORS

**Mark Budgeon, Maj, M.S.** *Mark Budgeon* is a Major in the U.S. Air Force Reserves, Lead Innovator at Air Force Global Strike Command, an Assistant Director of Operations at Barksdale AFB, LA, and an instructor/evaluator in the B-52H and E-11. He holds a Master of Science in Kinesiology from Penn State University and a Bachelor of Science in Mechanical Engineering from Penn State University.

**Brandon Wolf, Maj.** *Brandon Wolf* is a Major in the U.S. Air Force Reserves, Director of LRWrx and Assistant Director of Operations at Barksdale AFB, LA and an instructor/evaluator pilot in the B-52H Formal Training Unit. He holds a Master of Military Operational Art and Science from Air University and a Bachelor of Science in Systems Engineering, Human Systems from the United States Air Force Academy.

**Margaret Merkle, MBA, PMP.** *Margaret Merkle* is the Innovation Technology Chief, Simulators Division, Agile Combat Support Directorate, Air Force Life Cycle Management Center at Wright-Patterson AFB Ohio. She is responsible for leading the exploration, prototyping and integration of novel technologies into aircrew training and simulation systems, overseeing Pitch Day competitions and other innovation and research projects. She holds a Bachelor of Science in Business Administration in Management Information Systems and International Business from Bowling Green State University, and a Master of Business Administration from the University of Dayton.

**Donna Senft, PhD.** *Dr. Donna Cowell Senft* is the Chief Scientist for Air Force Global Strike Command. She serves as the AFGSC commander’s adviser on science, technology, technology transition and innovation on topics related to AFGSC core area programs. She holds a PhD in Materials Science & Engineering and a MS in Physics from the University of Illinois at Urbana-Champaign; and a BS in Engineering Science from Pennsylvania State University.

## INTRODUCTION

This paper compared the effectiveness of virtual reality (VR) training versus traditional training tools like: paper print-outs of the cockpit; life-like cardboard replicas of the cockpit; stationary simulators; and full-motion simulators. This study evaluated the use of VR training on the performance of checklist procedures, i.e. the act of moving switches in a particular sequence in order to properly operate an aircraft as published by the manufacturer. The findings of this study is the beginning of future studies to quantify performance outcomes in subject pilots trained by VR devices.

Since the first time mankind flew an aircraft, people have been developing ways to artificially re-create flight because of the associated costs with operating aircraft. Training in a simulated environment allows for more cost-efficient learning paradigms, while also providing the ability to present hazardous flight scenarios (e.g. an engine fire) without actually putting the student's life at risk. In the 1920's, Edwin Link understood the cost-benefit of a simulator, and developed one of the first well-documented trainers for flying called the "Link Trainer" (McIntosh, 1988).

As time passed, computers and technology advanced resulting in more sophisticated flight simulators. However, with an increase in fidelity came a short-term linear increase in cost followed by an exponential increase. The increase in simulator fidelity drove a cost-benefit challenge between achieving the desired training outcomes while doing so at a reasonable cost. Another way to describe this relationship is as follows: during lower stages of simulator fidelity, an increase in fidelity was met with a similar increase in cost; and when making the simulator as life-like as possible, any increase in fidelity is met with tremendous increases in cost. For example, compare a simulated cockpit where the switches move but there is no additional feedback, to a simulated cockpit where switches move and corresponding changes to lights, gauges and screens occur relatively identical to the actual aircraft. Simple tasks like checklist rehearsal can be practiced in the former example, whereas additional and more complex tasks (e.g. takeoff, or landing) can be accomplished on the latter; though at a significantly increased cost. There is an optimum fidelity-cost relationship as applied to a list of pilot training tasks, where more tasks can be trained as fidelity-cost increases, and before the relationship increases exponentially (Barnes, 1971; Rolfe, 1977; Rolfe and Caro, 1982; Rolfe and Staple, 1986; Finnegan, 1977).

The number of available training solutions has also multiplied as the technology has advanced (Flexman et al., 1972; Jacobs and Roscoe, 1975; Taylor et al., 2005a, 2005b). As new training devices are created, new human-machine interface dynamics are realized. Many studies have been done on the training efficacy and transfer of training suitability of ground-based aviation devices (Finnegan, 1977; Flexman et al., 1972; Jacobs and Roscoe, 1975; Roscoe and Williges, (1980); Kelly (1982); Kelly et al., (1985); Baldwin and Ford (1988); Taylor et al., 2005a). Most of these devices focus on the procedural nature, like checklist execution, but tend to neglect the physical and psychological components of the user experience (Swezey and Llaneras, 1997). With the relatively new technology of VR, there is potential to reasonably affect the physical and psychological domains with the inherent immersion that VR provides.

The introduction of VR in flight training simulations had several outcomes. Studies have shown that the impact of VR training can replace live-training (Champney et al., 2017), and can maximize training effectiveness with minimal cost (Padron et al., 2018). On the other hand, physiological maladaptations can lead to negative VR user experience, and overall negative training outcomes (Feng et al., 2017; Basak et al., 2019; Stanney, Lawson et al., 2020). Understanding the technological advancements of VR (Champney et al., 2017) with the physiological ramifications (Stanney, Hughes et al., 2021) should lead to a positive user experience. Combining the technological-physiological limitation understanding with a need to identify fidelity requirements in a training task (Muller et al., 2006) will lead to better development of VR training strategies (Stanney, Cohn et al., 2013). The aim of this paper is to include a quantitative analysis to the product of training strategy and identifying fidelity requirements keeping the physiological limitations in mind.

## METHODS

The purpose of this study was to investigate whether a different training technique – VR – improved the performance of student pilots above what could be expected using strictly traditional training techniques. Though a dynamic flight environment would be the ideal testing ground, the task of checklist training and execution was selected because of its simplicity. Checklist execution will also serve as a good starting point for quantitative analysis of VR as a new training tool. Because a laboratory was not available, an operational environment was used to perform the case study. In this environment, the subjects available for the study were limited to those of ground-school graduates and others in “casual status” waiting to start ground school. Though we desired to randomly select from a larger population to be able to extrapolate inferences on statistical conclusions, we decided to continue with a case study in order to develop test metrics for future studies.

All twelve of the subjects selected for the study had completed a year-long primary flight training program prior to arriving to the advanced aircraft training program. They had no other flight-related training prior to attending advanced aircraft training. Six of the subjects were graduates of a 3-month academic program for the advanced aircraft consisting of hundreds of hours of instruction, course study, and simulator training. They were randomly selected from a pool of twelve pilots that graduated the academic portion of the advanced aircraft training program. The six subjects selected for VR training were in “casual status” waiting to begin training for the advanced aircraft. They had received no other additional training since their primary training program. These six subjects were the only ones available for the VR-training portion of the study. No selection from a larger pool; hence, the limitation of this research to a “case study.”

The graduates of the academic program (no-VR) for the advanced aircraft were limited to traditional training tools. The traditional training tools consisted of the following: computer-based training in the form of presentations or written lectures; video lectures; in-person lectures; a simplified (but not very accurate) physical mock-up of the cockpit; a more physically accurate mock-up of the cockpit called a CPT (cockpit procedures trainer); and the most accurate aircraft representation called the WST (weapon-system trainer). The physical mock-up (fig. 1) has most of the switches found in the actual aircraft, but none of the switches would react as they would in the actual aircraft (e.g. turning “on” the battery in the aircraft would power on some gauges whereas nothing happens in the mock-up). The CPT (fig. 2) is another physical representation of the cockpit, but the switches feel and move like the actual aircraft, and also activate simulated systems (e.g. when the battery switch is turned “on” respective systems are powered). The “Weapons System Trainer” (WST) functions just like the CPT, but adds out-the-window visual simulations as well as full-motion. The ground-school graduates (no-VR) received training in the simplified mock-up, the CPT and the WST. They had zero training with the VR device.



**Figure 1:** Mock-up of the cockpit from the left seat.



**Figure 2:** Cockpit Procedures Trainer from the left seat

The six subjects that were exclusively VR-trained (yes-VR) had not started any formal advanced aircraft training, and had zero hours of experience in the advanced aircraft. Each subject in the yes-VR group was afforded 20 hours of time to train exclusively on checklist execution in VR. The VR experience provided a life-like cockpit visually where switches and gauges would react to user input, and the reactions were accurate to actual aircraft

designs. The subjects interfaced with the VR environment with two hand-held game controllers with joysticks and various buttons. These joysticks were the methods which the subject would interface with the VR cockpit. Specifically, the VR user would point to a switch, or control, in the VR cockpit and use the controller's joystick and buttons to manipulate the switch. Noise to simulate events like engine-start was embedded in the software in order to immerse the user. The cockpit was rendered by an artist to look as realistic as possible, and the cockpit components (switches, screens, etc) were constructed with high accuracy. This was designed such that the VR-user's movements would be as close to life-like as possible. A tutorial-like experience was embedded into the VR experience such that no human-guidance was needed to move from one checklist step to the next. An additional evaluation-like experience was embedded into the VR-training system so that the subject pilots could independently self-evaluate their performance.



**Figure 3:** Screenshot of VR device from left seat.



**Figure 4:** Actual aircraft from left seat.

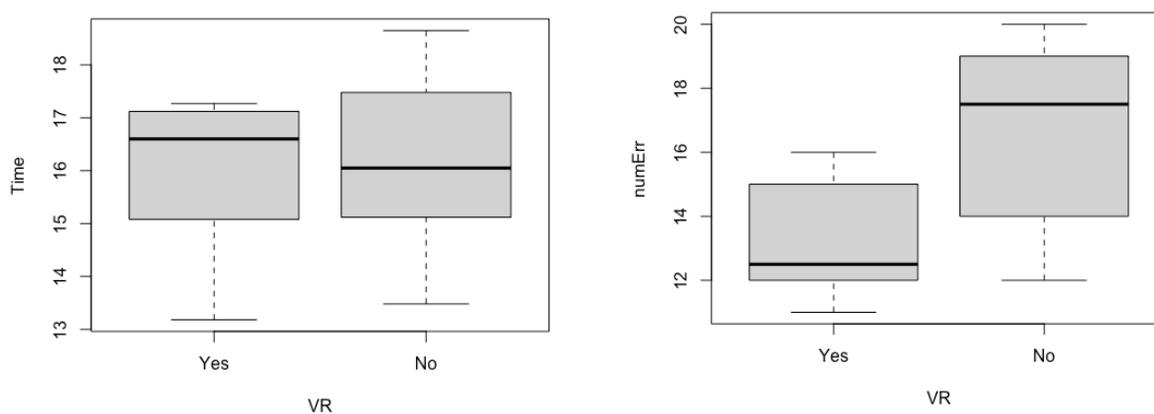
At the conclusion of the training, both populations were tested in a full-motion simulator called the WST. The WST provides the most life-like experience. All of the screens and switches are present and moveable switches provided resistance nearly identical to the real-life aircraft. Other haptic conditions like ambient noise and motion are present in the WST. The subjects were asked to complete the BEFORE STARTING ENGINES checklist, and checklist errors (numErr) and time-to-completion (time) were measured. There were 49 checklist steps for the observers to monitor and determine whether the subject did the step correctly, or not. An instructor was placed in the other crew position to run the checklist. Two observers were present to measure time to completion and checklist errors. A statistical analysis was performed on one factor (training type) with two levels (no-VR and yes-VR), and two outcome variables: number of errors (numErr), and time to complete the task (time). The tested hypothesis was whether the yes-VR subjects made statistically fewer errors, and/or completed the checklist faster than the traditional-trained (no-VR) subjects. The error location (checklist step) and type of error (omitted, skipped but later corrected, incorrectly performed) was also recorded for further analysis, but were not relevant to the hypothesis.

Some assumptions were made for this study: 1) The no-VR group was randomly selected from a group whereas the yes-VR group consisted of all available casual subjects. First, this was the nature of the operational environment that we recognized, and one of the leading factors that limits this research to a case-study. 2) Training provided in a VR environment is of substantially different quality than traditional training methods. In other words, we assumed that one hour of training in a VR device is of better training quality than one hour in traditional devices. Ideally, we wanted to put both groups through the ground-school of the advanced aircraft where one group only uses the traditional tools, and the other uses the VR device in place of the traditional tools. The operational environment limited the scope of this research such that this comparison was accomplished between academic graduates (no-VR) and those awaiting training (yes-VR). However, we believed that by comparing academic graduates versus those awaiting training would be a powerful indicator for future studies. When simply observing the number of hours of training each group received, the VR-trained subjects had less than 95% training hours than the non-VR trained subjects. 3) How to test the performance of both groups? We selected the WST because it represents a physical

device that closely represents the actual aircraft, and we were interested in performance improvements with operating the actual aircraft. We selected the BEFORE STARTING ENGINES checklist because that was a complicated, difficult checklist (where errors are commonly made) that contains switches present across all of the training devices. 4) The instructor in the opposing seat performed the same for each subject during the test. There were two instructors used for test, and both had thousands of hours of flight experience as well as being Formal Training Unit instructors. We assumed that the different instructors would have little impact on the subject's performance, and was not included as a variable in the model.

## RESULTS

The raw data from the study included completion time and number of errors for each subject. Box plots (*fig.5*) of the data show the upper/lower bounds of one standard deviation, the upper/lower bounds of the data set, and the mean. In order to ensure the data was suitable for ANOVA, MANOVA, and Cohen's D measures, several tests were performed: Shapiro-Wilk Normality Test, testing the positive determinant of the variance-covariance matrix; Box's M-test for Homogeneity of Covariance Matrix; and Hotelling's t-squared test. The results of the Shapiro-Wilk test (p-value = 0.1082) gave a p-value greater than 0.05, which informs that there is no statistically-significant abnormal distribution of the variable, and we could accept that the two variables have a normal distribution. The determinant of the variance-covariance matrix is 24.11571, indicating that the data are positively proportional and generally not dispersed. Box's M-test for Homogeneity of covariance matrices between yes-VR and no-VR factors reveal a p-value of 0.4958, where a value greater than 0.05 shows that there is no statistically significant difference between the yes-VR variance-covariance matrix and the no-VR variance-covariance matrix. Hotelling's t-squared test reveals that the grouping of the two means (time, number of errors) under the yes-VR factor is not statistically different from the grouping of the two means under the no-VR factor. For subjects trained with VR, the average time to completion was 15.98 min with one standard deviation of 1.60 min. The average number of errors for VR-trained subjects was 13.17 with one standard deviation of 1.94. For subjects trained without VR, the average time to checklist completion was 16.14 min with one standard deviation of 1.80 min. The average number or errors was 16.67 with one standard deviation of 3.08. Results of the MANOVA reveal that there is no statistically significant difference (p-value = 0.1354) in the mean of the no-VR time\*numErr interaction and the mean of the yes-VR time\*numErr interaction. The ANOVA analysis shows that there was no statistically significant difference in the means of task time-to-completion (time p-value = 0.8713), but there was a statistically significant difference in means for the number of errors (numErr p-value = 0.0402) between yes-VR and no-VR subjects. Yes-VR subjects made statistically significant fewer errors than no-VR subjects.



**Figure 5:** Box plots of the “time” and “numErr” variables against the Factors “yes-VR and “no-VR”. The dark line is the median. The mean is the center of the box plot.

## DISCUSSION

**Quantitative VR investigation.** When we first started working with VR as a training tool, we knew that we needed to evaluate whether VR improved the performance outcomes for the subject pilots. We also knew that we had limited resources – we live in an operational environment and not a lab; the subject pilots are limited to a prescribed training timeline restricting *when* we would have subjects to do the study; the number of subjects *per class* are limited; and the number of subjects on casual status were limited. We also understood that VR alone would not train, but would require an instructional-like environment embedded into the user experience. A lot of time was spent on developing the user experience so that educational outcomes of checklist usage were likely to happen within the user experience itself. It was easy to become distracted in the study of this haptic experience alone. So, given this situation, we decided to focus on the training outcomes of the traditional versus VR with the resources we had at the time in order to inform future studies. Simultaneously, we decided to minimize the number of variables in the model while making reasonable assumptions. This minimization of variables was one of the reasons why we decided to observe a static task, like checklist execution, instead of a more complex task like flying. We concluded that time-to-completion (time) and number of errors (numErr) would be sufficient measures of performance given the situation regarding resources and assumptions.

**Training in VR with “guidance” software reduces error rate.** The model we used to test the hypothesis included the variables of time, numErr, and the interaction of time and numErr (time\*numErr). The two factors were training without VR (no-VR) and training with VR (yes-VR). The analysis showed that the only variable where the means from the two groups differed significantly was numErr. Two ways to explore this outcome are: 1) the subjects from both groups completed the task in the same amount of time; and 2) the subjects that trained with VR made fewer errors than those that trained under traditional methods. The VR trained subjects made an average of three fewer errors (p-value = 0.04), or roughly 19% less, than the traditionally trained subjects. Recall that the traditional subjects in this study received at least ten times more hours of total training, with twice as many hours at the seat of some type of training device. The reduction in error rate is likely attributed to haptic (human-machine interface), or user experience, that describes the VR trainer in its entirety: it is immersive, fun, realistic, and displays relevant kernels of information appropriate to task at the right time directly in the user’s field of view. It is this combination of visual information delivery, realism, and timing of delivery while also being enjoyable for the user that makes the VR experience unique. Additionally, there is a computer monitoring the user’s VR experience and “tracking” their progress to ensure they correctly complete the step before moving on.

Consider the static “cardboard” trainers that the non-VR trained subjects used for their training. The switches exist, but not all of them are displayed like the user would expect in the real aircraft. The mock-up doesn’t respond to switch interaction from the user (i.e. no lights illuminate or gauges move) because there are no sub-systems behind the switches and gauges. The user is forced to imagine how the cockpit environment would respond after a switch interaction, or interactions. Further, the user of the cardboard mock-up depends on a human-observer to catch mistakes, or their own ability to monitor their performance while doing the task. In either case, the user has to synthesize learning how to correctly execute a checklist in the right order, imagine how the cockpit environment responds, further imagine how to correctly analyze the cockpit response, move on to the next step, repeat the switch-environment feedback analysis, and evaluate whether this process was done correctly. This is a complex and intensive human performance process.

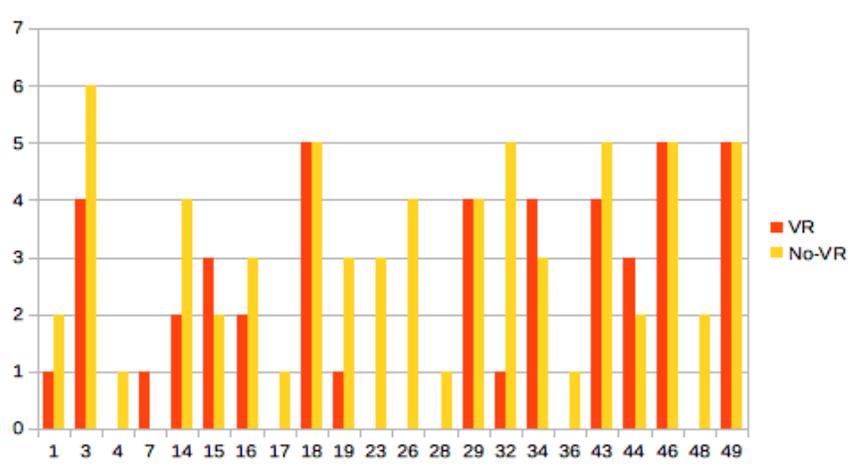
In the traditional training methods, the observer has to recognize that the subject using the training device has performed a step incorrectly *and* give the appropriate guidance to correct errors. The guidance employed by the VR training device is such that the appropriate item is highlighted, and the highlight does not disappear until the user has done the checklist step correctly. The guidance method employed in the VR device performs better than a human observer. When a VR user desires to test their skill, they can exit this mode and have the VR device evaluate their checklist performance. Sensors like eye-tracking improve the performance of the guidance model in VR by employing simple algorithms to determine whether the user is looking in the right place, and for the right amount of time.

**Proprioception \* VR environmental response > virtual feel of the switch.** A common critique of VR cockpit training environments is that the environment with which the user interacts is virtual. The VR user does not actually feel the weight of the switch being flipped, or the resistance of a dial being turned, and this lack of tactile

feedback detracts from human performance. Typically critics use the terminology “muscle memory” to describe their critique. The VR device used in this study employed the use of joysticks for the human-virtual environment interaction. The users would press a button with their middle finger, or ring finger, or pivot a joystick with their thumb to make the switches in the VR environment move in the user’s desired direction. In other words, the VR user had to learn an extra skill to be able to interact with the VR environment. Despite this, they made fewer checklist errors than non-VR users in the actual test environment. This suggests that it is not the feel of the switch manipulation that is important, but the combination of proprioception and the visual reception of environmental responses (e.g. what lights illuminate after the battery switch is flipped on, and where they occur). This finding supports other research that visual and proprioceptive coupling support spatial training transfer more accurately than visual systems alone (Stanney, Cohn, et al., 2013). In the case of checklist learning and execution, proprioception and environmental response were important. However, actual switch “feel” and other haptic phenomena may be crucial for more dynamic tasks (e.g. yoke feedback during air refueling).

**Deep dive in error location: a result of identifying fidelity requirements with a VR training strategy.**

After discovering that the error rate decreased for subjects trained in VR environments, we decided to look deeper into what errors the two groups made. There were a total of 49 checklist steps that a subject could perform in the test. Of those 49 checklist steps, the yes-VR subjects made an error on only 15 of those steps, whereas the traditionally trained subjects made errors on 21 steps. A histogram was plotted to show these results (fig.6). Only one checklist step was unique to the yes-VR subjects’, step 7. One yes-VR subject failed to look at the MAIN TANK low warning light. This only happened once across the entire group of yes-VR subjects. Traditionally trained subjects (no-VR) made 6 unique checklist step errors (steps 4, 17, 23, 26, 28, and 48) – that is 5 more unique checklist errors than yes-VR subjects. We also looked at checklist steps that had higher frequency of errors (i.e. 4 or more errors) for VR subjects and found that 6 out of 7 were also higher frequency for non-VR subjects. These checklist steps (3, 18, 29, 34, 43, 46 and 49) were some of the more complicated steps that required several sub-steps for successful completion. What is more interesting is that only 3% of the errors yes-VR subjects made were unique, whereas approximately 20% of the errors made by traditionally trained subjects were unique. This suggests that the VR training technique was superior to traditional training techniques as a result of the planning for fidelity requirements (Muller et al., 2006) and a strategy to incorporate VR into checklist execution training (Stanney, Cohn et al., 2013).



**Figure 6:** Histogram of yes-VR and no-VR error locations. X-axis is error location, y-axis is frequency.

## CONCLUSIONS

We originally set out to determine whether this VR training tool would be an effective investment from the stand point of improving the performance outcomes of the subject pilots. Our hypothesis was that a VR training environment, coupled with software designed to effectively guide the subject through a task, would improve the performance outcomes of a USAF pilot measured by a reduction in time to complete a task or reduction in errors. Due to limitations in resources, working in an operational environment, and shorter timeline, we limited the variables to determine performance outcomes to time-to-completion (time) and number of errors (numErr). The study consisted of 12 subjects, 6 of them were academic graduates and 6 were casual subjects, who performed the BEFORE STARTING ENGINES CHECKLIST in the WST. The yes-VR subjects completed the checklist in the same amount of time as the academic graduates. However, the yes-VR subjects made significantly fewer errors ( $p$ -value = 0.04) than the academic graduates even though they had never touched the aircraft prior to this study. We attribute the improvement in performance to: 1) the guidance model built into the VR training device; and 2) the combination of proprioception with environmental responses to user inputs. The results of this analysis agree with other studies that, when task-fidelity and VR strategy are planned for (Muller et al., 2006; Stanney, Cohn et al., 2013; Stanney, Hughes et al., 2021), VR can maximize training effectiveness while minimizing cost (Padron et al., 2018; Stanney, Lawson et al., 2020).

Due to the nature of the case-study, the next step is to perform a similar study where the yes-VR subjects conduct the ground-school program using the VR training device, and compare that group with traditionally-trained pilot groups. This would give us greater insight into the performance outcomes, and allow for generalized findings for VR training in checklist execution. Additionally, looking at the training efficacy and transfer-of-training *while* the VR device is being used would give us greater understanding about variables to observe during a training session, or across several training sessions. The authors hypothesize that the set of variables for performance outcomes will have members exclusive to the set of variables relative to monitoring transfer-of-training and training efficacy. Further, the set of variables important for measuring the performance outcomes for a static task like checklist execution, will be different than a set of variables for a dynamic task (e.g. taking off, landing, etc). Finally, we should investigate the relationship of variables between performance outcomes and those relative to training efficacy for dynamic flight tasks.

## ACKNOWLEDGMENTS

The authors would like to thank Col William Fish and Col David Anderson for the opportunity to perform this study during the busy instruction season; Maj Justin Stephenson and Maj Justin Cedarleaf for their work organizing the test and collecting data; and Dr. Kay Stanney for supplying some of the articles for our literature review.

### **Non-Endorsement Disclaimer:**

The views, opinions and/or findings contained in this research/presentation/publication are those of the authors and do not necessarily reflect the views of the Department of Defense and should not be construed as an official DOD/Air Force position, policy or decision unless so designated by other documentation. No official endorsement should be made. Reference herein to any specific commercial products, process, or service by trade, name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government.

## REFERENCES

- Baldwin, T.T. and Ford, J.K. (1988) Transfer of training: a review and directions for future research, *Personnel Psychology*, 41(1), 63–105.
- Barnes, A.G. (1971) *Objectives of Simulation*. Technical Report, AGARD-CP-79-70, January 1971.

- Basak, U., Kazempourradi, S., Yilmaz, C., Ulusoy, E., & Urey, H. (2019). Dual focal plane augmented reality interactive display with gaze-tracker. *OSA Continuum*, 2(5), 1734-1745.
- Carroll, M., Champney, R., Wang, W., Woodman, M., Kokini, C., Haddad, D., & Hale, K. (2012) *Return on Investment Tool for Effective Medical Training (RITE-MT)*, Final Report N00014-12-M-0105, Design Interactive Inc., Oviedo, FL.
- Champney, R.K., Stanney, K.M., Milham, L., Carroll, M.B. & Cohn, J.V. (2017) An examination of virtual environment training fidelity on training effectiveness. *Int. J. of Learning Technology*, 12(1), 42-65.
- Fang, W., Zheng, L., Deng, H., & Zhang, H. (2017). Real-time motion tracking for mobile augmented/virtual reality using adaptive visual-inertial fusion. *Sensors*, 17(5), 1037.
- Finnegan, J.P. (1977) *Evaluation of the Transfer and Cost Effectiveness of a Complex Computer-Assisted Flight Procedures Trainer*, Technical Report ARL-77-07/AFOSR-77-6, University of Illinois, Aviation Research Lab, Savor, IL.
- Flexman, R.E., Roscoe, S.N., Williams Jr., A.C. & Williges, B.E. (1972) "Studies in pilot training: the anatomy of transfer", *Aviation Research Monographs*, 2(1), 1-87.
- Jacobs, R.S. and Roscoe, S.N. (1975) *Simulator Cockpit Motion and the Transfer of Initial Flight Training*, Technical Report ARL-75-18/AFOST-75-8, University of Illinois, Aviation Research Lab, Savoy, IL.
- Kelly, A.I., Orgel, R.F. and Baer, D.M. (1985) Seven strategies that guarantee training transfer, *Training & Development Journal*, 39(11), 78-82.
- Kelly, H.B. (1982) A primer on transfer of training, *Training & Development Journal*, 36(11), 102-106.
- McIntosh, D.M. (1988) The Evolution of Instrument Flying in the US Army. *Air Command and Staff College*, Maxwell AFB, AL. April 1988 ACSC-88-1760
- Muller, P., Cohn, J., Schmorow, D., Stripling, R., Stanney, K.M., Milham, L., Whitton, M.C., & Fowlkes, J.E. (2006) The Fidelity Matrix: Mapping System Fidelity to Training Outcome. *IITSEC 2006*, Paper No. 2994.
- Padron, C., Mishler, A., Fidopiastis, C., Stanney, K.M., Fragomeni, G. (2018) Maximizing Return on Training Investment in Mixed Reality Systems. *IITSEC 2018*, Paper No. 18075.
- Rolfe, J.M. (1977). Keeping up on the Ground. *The Aeronautical Journal*, 81(799), 281-292.
- Rolfe, J.M., Caro, P.W. (1982) Determining the Training Effectiveness of Flight Simulators: Some Basic Issues and Practical Developments. *Applied Ergonomics*, 13(4), 243-250.
- Rolfe, J.M., Staples, K.J. (1986) *Flight Simulation*. Cambridge University Press. ISBN 0 521 35751 9
- Roscoe, S.N. and Williges, B.H. (1980) Chapter 16: measurement of transfer of training, in Roscoe, S.N. (Ed.): *Aviation Psychology*, pp.182-193, Iowa State University Press, Ames.
- Stanney, K.M. Cohn, J., Milham, L., Hale, K. Darken, R., & Sullivan, J. (2013) Deriving Training Strategies for Spatial Knowledge Acquisition from Behavioral, Cognitive, and Neural Foundations. *Military Psychology*, 25(3), 191-205.
- Stanney, K.M., Hughes, C., Fidopiastis, C., Jasper, A. (2021) Augmented Reality in Tactical Combat Casualty Care: Physiological Ramifications. *IITSEC 2021*, Paper No. 21136.

Stanney, K.M., Lawson, B.D., Rokers, B., Dennison, M., Fidopiastis, C., Stoffregen, T., Weech, S., & Fulvio, J.M. (2020). Identifying causes of and solutions for cybersickness in immersive technology: Reformulation of a research and development agenda. *International Journal of Human-Computer Interaction*, 36(19), 1783- 1803.

Swezey, R.W. and Llaneras, R.E. (1997) 'Models in training and instruction', in Salvendy, G. (Ed.): *Handbook of Human Factors and Ergonomics*, pp.514–577, Wiley-Interscience, New York, NY.

Taylor, H.L., Talleur, D.A., Emanuel, T.W. & Rantanen, E.M. (2005a) *Effectiveness of Flight Training Devices Used for Instrument Training*, Final Technical Report AHFD-05-9/FAA-05-4 [online] <http://www.humanfactors.uiuc.edu/Reports&PapersPDFs/TechReport/05-9.pdf> (accessed 7Aug2005).

Taylor, H.L., Talleur, D.A., Rantanen, E.M. and Emanuel, T.W. (2005b) 'The effectiveness of a personal computer aviation training device (PCATD), a flight training device (FTD), and an airplane in conducting instrument proficiency checks', paper presented at the *13th International Symposium on Aviation Psychology*, Dayton, OH.