

# Training Effectiveness and Validation of a VR HMD-based Simulator for Air Force Pilots: Preliminary Findings

Ramy Kirollos and Wasim Merchant

Defence Research and Development Canada - Toronto  
Research Centre

Toronto, ON, Canada

[ramy.kirollos@drdc-rddc.gc.ca](mailto:ramy.kirollos@drdc-rddc.gc.ca),  
[wasim.merchant@drdc-rddc.gc.ca](mailto:wasim.merchant@drdc-rddc.gc.ca)

## ABSTRACT

There is great interest in extended reality (xR) head-mounted displays (HMDs) for military training. But not enough empirical research has been performed on training effectiveness and validation of these devices in the military training domain. The current study investigated the training effectiveness and validation of a virtual reality (VR) HMD-based simulator in Royal Canadian Air Force (RCAF) pilot training. We compared Novice and Expert Pilots on the Overhead Break Flight Manoeuvre in a VR HMD-based simulator. Evaluators assigned a score from 1-5 on a validated criterion used by the RCAF on 7 Overhead Break attempts made by each participant. We hypothesized that Expert pilots would perform significantly better than Novice pilots, demonstrating that Expert Pilots find the VR HMD-based trainer similar to live flying of the aircraft represented in VR, in part validating the VR HMD-based trainer. Findings supported this hypothesis. Importantly, Novice Pilots improved significantly on the Overhead Break over the course of the experiment demonstrating the training effectiveness of the VR HMD-based trainer. Cybersickness scores obtained from participants during the 45-min experiment were negligible. Taken together, these preliminary findings indicate that the VR HMD-based trainer evaluated herein is a valid and effective method for training the Overhead Break. Future research should aim to validate VR HMD-based simulators as done in the present study and investigate if their use significantly improves student pilot performance in live flight. This will broaden our understanding for the successful and safe integration of VR HMD-based simulators in military flight training and other military occupation training applications.

**Keywords:** VR HMD; training effectiveness; reverse transfer of training; pilot training; flight simulation

## ABOUT THE AUTHORS

**Dr. Ramy Kirollos** is a Defence Scientist at Defence Research & Development Canada and holds an Adjunct Assistant Professorship in the Dept. of Kinesiology & Health Sciences at the University of Waterloo. He earned his PhD from Carleton University in Cognitive Science. His research is in human vision, self-motion perception, neurophysiology, visual-vestibular sensory integration, depth perception, vehicle simulation, extended reality technologies and training. He develops new simulation and training solutions that support the Canadian Armed Forces.

**Wasim Merchant** is a Research Technologist at Defence Research & Development Canada in the Human Effectiveness Section. He has a BSc. in Kinesiology from the University of Waterloo (2019). Wasim's research interests are in the realm of visual perception using extended reality technologies, but he has also contributed to research in learning & training and mental health. He has extensive experience conducting experiments directly involving members of the Canadian Armed Forces.

# Training Effectiveness and Validation of a VR HMD-based Simulator for Air Force Pilots: Preliminary Findings

Ramy Kirolos and Wasim Merchant

Defence Research and Development Canada - Toronto  
Research Centre

Toronto, ON, Canada

[ramy.kirolos@drdc-rddc.gc.ca](mailto:ramy.kirolos@drdc-rddc.gc.ca),  
[wasim.merchant@drdc-rddc.gc.ca](mailto:wasim.merchant@drdc-rddc.gc.ca)

## INTRODUCTION

Over the past decade, commercial off the shelf extended reality (xR) head-mounted displays (HMDs) have become popular and accessible in gaming, entertainment, the fine arts, tourism, construction, real estate, research and development, military, aviation, education, medicine and more. It is important to ensure that these systems are valid and effective training and operational tools in safety-critical domains such as military training and operations. This study investigated the training effectiveness (TE) and validation of a virtual reality (VR) HMD-based flight simulator with the aim to better understand if VR HMD-based flight simulators can be used constructively in the military pilot training pipeline because there is little available empirical research investigating this topic.

### Overview of Empirical Pilot Training Effectiveness Research using xR HMDs

Research by the United States Air Force (USAF) presents a small body of publicly available research on testing xR HMDs for pilot TE (McCoy-Fisher et al., 2019; Mishler et al., 2022; Severe-Valsaint et al., 2022). These studies are part of a formal VR HMD-based training pipeline for ab initio pilot training previously named Pilot Training Next for US Forces Aviators and has since been renamed 'Undergraduate Pilot Training (UPT)'. Mishler et al. (2022) evaluated Project Avenger - a flexible competency-based learning approach focusing on the use of VR, compared to legacy fixed-wing US Forces pilot training syllabi. Mishler et al. evaluated pilot grades in the Avenger program compared to legacy training. They found that Avenger program student pilot grades were lower than archival student pilot scores that did not complete the Avenger program but that time to train pilots was significantly shorter in the Avenger project compared to legacy training. Though, cohorts in the Avenger program condition were smaller, possibly explaining this finding.

Severe-Valsaint et al. (2022) measured task performance scores and number of task repeats as objective measures when evaluating the use of an imbedded intelligent tutoring system in a VR trainer called 'VIPER' with US Navy pilots. They examined flight test performance grades and retries after rehearsal in 4 conditions: 1) archival data with no VR training intervention; 2) unstructured VR use; 3) structured VR use; and 4) VIPER intelligent tutor-aided training. They found significant performance increases in evaluations in 3 conditions relative to the archival control condition. Findings from Severe-Valsaint et al., in contrast to those in Mishler et al. suggest that VR-based training provides an advantage over no training (e.g. archival cohort). However, authors reported that though VIPER shows promise, it was not yet ready for use within the UPT syllabus.

Munn, Jarmasz and Deluce (2023) compared the use of a Mixed Reality (MR) HMD to two types of 2D displays in a fast jet formation flight task by examining pilot distance from the lead jet. Munn et al. found no significant difference in separation distance across all three visual displays. However, the MR HMD condition produced the largest separation distance, though not significant and the largest workload measured with the NASA TLX.

Cain, Magee and Kersten (2012) evaluated the TE of a helicopter VR HMD-based trainer for landing on a ship deck in Royal Canadian Air Force (RCAF) helicopter pilots. They tested novice and expert pilots and measured % correct of verbal commands + manual actions, simulator sickness and cognitive workload with the NASA TLX. They used a reverse transfer of training (RToT) research design (AGARD, 1980). In a RToT experiment, it is hypothesized that if the simulator is a valid training aid, experts should perform better than novices in the experimental task. In Cain, Magee and Kersten (2012) Expert pilots performing well in the simulator, especially in early trials, would indicate that the simulator resembles the live task and would provide evidence validating the simulator. Cain et al. found this to be the case based on % correct data.

### **Brief Overview of Relevant Training Effectiveness Methods**

A traditional Transfer of Training (ToT) research design employs an experimental training intervention and is compared to the traditional training intervention (AGARD, 1980; Korteling, Oprins, & Kallen, 2013). However, in cases in which the experimental training intervention unknowingly produces negative transfer of training, participants may perform poorly in the traditional trainer, imposing a risk that training can harm user performance during operations. There are resultantly few TE experiments in safety-critical domains such as pilot training. This is likely in part because of the risk posed by experimental training interventions such as ToT designs because they are too reliant on the experimental training intervention (Allsop, Olonilua, & Yeo, 2021).

In contrast, RToT experiments allow for the testing and validation of a new training intervention without interfering with an established training method. Therefore, reducing the reliance on the training intervention and reducing the risk that it may unknowingly pose (AGARD, 1980). RToT experiments largely negate the risk of negative transfer because they are tested on an expert group. If expert group performance is high, this partly validates the training intervention (AGARD, 1980; Cain, Magee, & Kersten, 2012; Korteling, Oprins, & Kallen, 2013).

However, RToT designs have some drawbacks (AGARD, 1980). For instance, it may be the case that experts have enough generalized knowledge that they can easily adapt to the experimental training intervention. Additionally, the experimental training intervention may be too advanced for novices, perhaps making it ill-suited in contexts comparing novice and expert performance.

### **Present Study**

Building on the body of empirical research on pilot TE using xR HMDs presented above, in the present study, we aimed to assess the validity of VR HMD-based trainers for the RCAF to make informed decisions about the employment and integration of xR HMD-based training for RCAF student pilots. We performed an RToT study comparing Novice Pilots to Expert Pilots in a VR HMD-based trainer. We assessed performance of the Overhead Break (OHB) (i.e., Overhead Manoeuvre and/or Approach) over 7 repeats in both groups graded by an (instructor pilot) IP. We hypothesized that Expert Pilots would reach a high level of proficiency early in the experiment if the VR HMD-based trainer is a valid training aid, in line with hypotheses supporting RToT studies. We also hypothesized that if Novice Pilot Flight Proficiency Scores (FPS) increased significantly throughout the experimental session, this would indicate evidence supporting VR-HMD based TE for student pilots rehearsing the OHB when other training (e.g. Flight training devices, live flight) is not accessible. We measured participant cybersickness, Expert-rated fidelity of the VR HMD-based trainer relative to live flight of the modeled aircraft, and demographic data on flight hours and VR experience for additional comparisons. The experiment took place at 2 Canadian Forces Flight Training School (2CFFTS), in Moose Jaw, Saskatchewan, Canada where RCAF student pilots undergo Phase 2 pilot training.

In the present study, we have taken two measures to mitigate potential risks posed by RToT designs: 1) We obtained Simulator Fidelity (SiFi) scores using a scale developed by Bush, Lamb and Braun (2022) from Expert Pilots to measure their perception of how closely the VR HMD-based trainer matches the live aircraft it represents. SiFi scores were intended to help contextualize Expert FPS and to elucidate the relationship between simulator fidelity and flight performance on the OHB. 2) Measures of performance over time were taken to determine if Novice Pilots improved in the experimental task over the course of the experiment.

### **METHOD**

**Study Design**

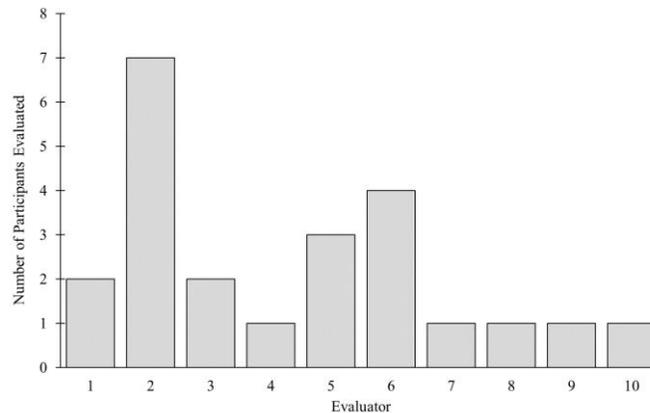
A mixed model design with condition (Expert v. Novice) as the between-subjects independent variable, and repetition (7 repetitions of the OHB) as the within-subjects independent variable was used in this experiment. All participants in both conditions performed the same experimental task.

**Participants**

All study participants were volunteers from the RCAF pilot community at 2CFFTS. Expert Pilot participants were qualified instructor pilots (IPs) that had flown the Harvard II aircraft (i.e. the aircraft represented in simulation) within the last 6 months. Participants in this group were required to have a minimum of 100 hours each of live flight in the Harvard II. Novice Pilots in this experiment were defined as student pilots that completed Phase 1 pilot training (i.e. ground school) with up to 5 hours of simulator experience. Novice participants had no experience flying the Harvard II aircraft.

Twenty-three participants volunteered in the experiment. Fourteen participants met the criteria to be in the Expert group ( $M = 37.3, SD = 9.1$  yrs.). The average Harvard II flight hours for participants in the Expert group was 989.3 hours ( $SD = 1227.0$ ) per participant. Nine participants were in the Novice group ( $M = 29.8, SD = 6.5$  yrs.).

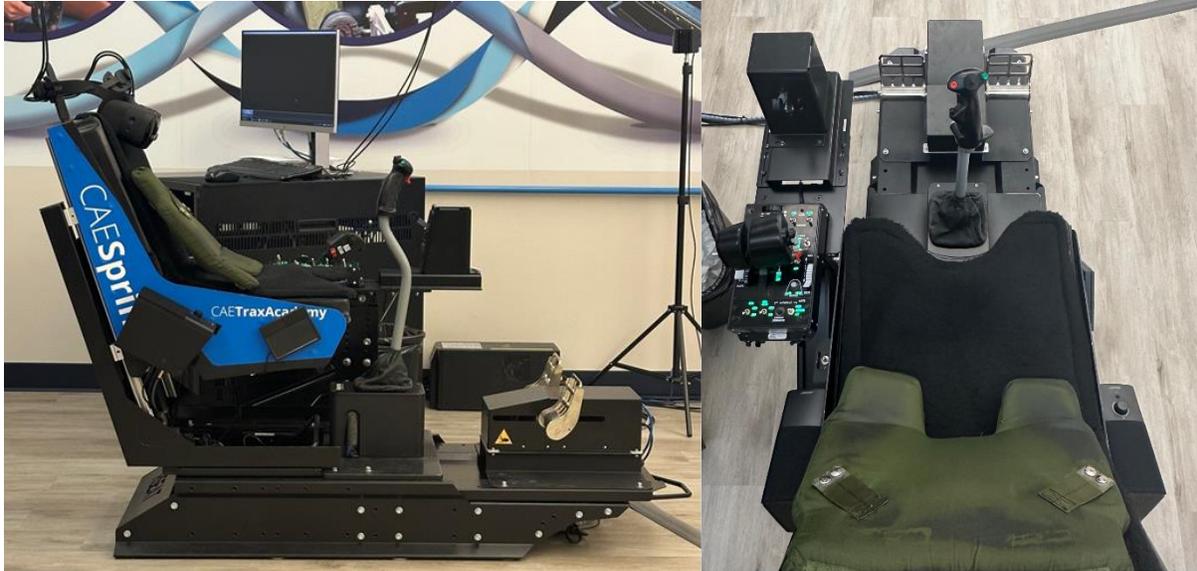
10 IPs also participated as evaluators for the experiment and were assigned to evaluate participants in the Novice and Expert groups. The distribution of frequency of each evaluator in an experimental session is displayed in Figure 1. Evaluators were assigned to a participant based on their availability a day or less prior to the scheduled experimental session. Every effort was made to distribute evaluators evenly across participants.



**Figure 1. Distribution of evaluators used for experimentation. Ten qualified IPs volunteered to evaluate Experts and Novices.**

**Materials and Stimuli**

The CAE Sprint Chair in Figure 2 is a VR HMD-based flight simulator used in this experiment. It provides a visual virtual environment and physical flight controls (e.g., throttle, flap switch, yoke, etc.). The virtual environment in the Sprint Chair is displayed through the Varjo VR 3 HMD. The Varjo VR 3 is tethered to a PC and has a 2880 x 2720-pixel resolution per eye and a 115° horizontal and a 90° vertical combined (i.e. binocular viewing) native field of view (Varjo Tech, 2020). The CAE Sprint Chair affords users haptic feedback on the yoke and rudder pedals, and audio through two speakers mounted on the left and right sides of the Sprint Chair.



**Figure 2. Side-view of the Sprint Chair with the Varjo VR 3 HMD (left). Top-down view of the Sprint Chair (right).**

The Sprint Chair employs CAE proprietary software and hardware components that replicate the physical Harvard II aircraft that is used for Phase 2 Pilot training at 2CFFTS pictured in Figure 3. The CAE Medallion Image Generator provides out-the-window graphics. The CAE Simfinity Virtual Cockpit provides high-fidelity graphics of the inside of the Harvard II aircraft (i.e., cockpit) and employs Brunner Flight Controls, and the EDSI T-6 Throttle (Thrustmaster A-10C Throttle). The A-10C throttle was used in place of the Harvard II Throttle because a functioning Harvard II throttle was not available at the time of the experiment. This remained consistent for all participants throughout data collection and was made known to all participants during initial familiarization period in the experiment. For all experimental trials, clear, sunny sky daytime conditions over a virtual rendition of CFB Moose Jaw were used.

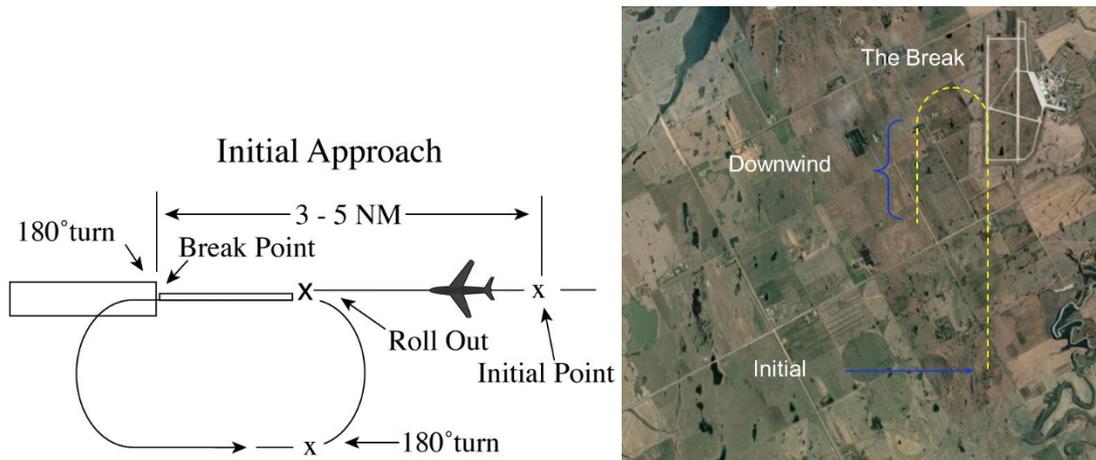


**Figure 3. The CT-156 Harvard II trainer aircraft replicated in the CAE Sprint Chair (Royal Canadian Air Force, 2020).**

### **Task: Overhead Break**

The OHB is a procedure employed by military pilots to improve the efficiency of landing without requiring the aircraft to fly at low speeds for a long time prior to landing. It consists of two 180° energy-depleting turns that cause a rapid

reduction in airspeed in preparation for landing (FAA, 2021). Figure 4 (left) OHB procedure is in Figure 4 (Left). Novice participants received an introductory in-class brief on the OHB manoeuvre via a slide deck one-week prior to the experiment and a short refresher brief by the evaluator at the start of the experiment.



**Figure 4. Typical OHB procedure (left; FAA, 2021). Representation of the experimental protocol for the OHB. The experiment begins at initial and is complete after the downwind phase (right).**

The current experiment consisted of the initial, the first break, and downwind, meaning that only one 180° turn of the OHB. Participants were not required to conduct the final turn and landing as would typically be part of the OHB. The task participants were instructed to perform is summarized in Figure 4 (Right).

**Measures**

**Flying Proficiency Levels Criterion**

Grading of the OHB was completed by the assigned Evaluator after each repetition using the standard RCAF Proficiency Levels criterion (Royal Canadian Air Force, 2014). This Flying Proficiency Levels criterion consists of five levels ranging from “Level 5 – The candidate completed the task without assistance and without error” to “Level 1 – The candidate was not capable of completing the task...”. Performance was assessed based on the specific assessment criteria of the OHB training plan summarized in Table 1. Scores administered by the Evaluator are referred to as ‘Flight Proficiency Scores’ (FPS) herein given for the entire maneuver.

**Table 1. Assessment criteria of the overhead break training plan (Royal Canadian Air Force, 2014).**

Flight Manoeuvre	Phases	Assessment Criteria
<b>Overhead Break (OHB)</b>	Initial Point	Lookout – Traffic Awareness Wind Assessment Radio/Telephone Communication Accuracy
	Break	Lookout Wind Assessment Altitude/Bank Angle and Airspeed Control Accuracy Effective use of APT/PAT Procedures
	Downwind	Lookout Wind Assessment Gear and Flap Selection Checks

		Altitude and Airspeed Control Effective use of APT/PAT Accuracy
--	--	---

**Simulator Sickness Questionnaire (SSQ) and Custom Demographic Questionnaire**

The SSQ (Kennedy et al., 1993) consists of 16 symptoms (e.g., nausea, fatigue, discomfort, etc.) that are rated by participants on a four-point Likert scale (*not at all* = 0, *mild* = 1, *moderate* = 2, and *severe* = 3). It is used to assess sickness state relevant to use of simulators and other environments where motion sickness can occur. The SSQ was used to index cybersickness before and immediately after the experiment. Participants’ age, sex, flying experience, familiarity with VR and other demographic metrics were recorded in a demographic questionnaire we developed.

**Simulator Fidelity (SiFi) Scale**

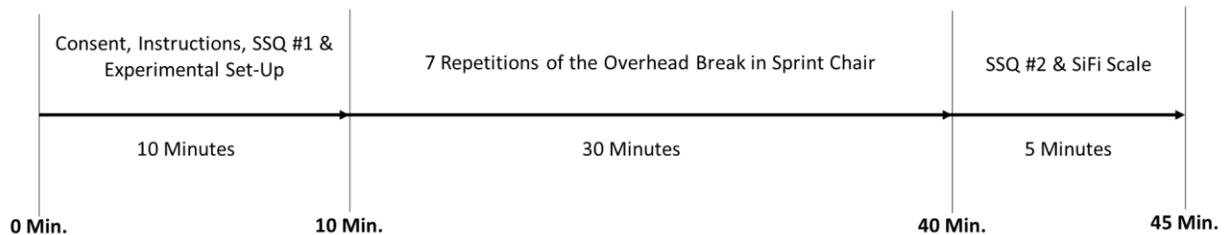
Expert participants provided subjective ratings of the Sprint Chair using the SiFi Scale (Bush, Lamb, & Braun, 2022). Expert participants were asked to rate the Sprint Chair’s Physical Fidelity (i.e., visual, auditory, tactile) and Cognitive Fidelity (i.e., user interaction, system behaviour) relative to the actual Harvard II aircraft on a 0 – 5 scale. A rating of 5 on the scale indicates that the simulator is indistinguishable from the real-world Harvard II aircraft in that category, while a rating of 0 indicates that an element present in the real aircraft is not present in the simulator.

**Procedure**

Study candidates were recruited from the pilot community at 2CFFTS through emails, posters and class announcements. Participants provided written informed consent for their voluntary participation in the experiment approved by the DRDC Human Research Ethics Board (Protocol #2023-23). Participants completed the demographic questionnaire and SSQ. They were then seated in the Sprint Chair and donned the Varjo VR3 HMD. Physical seat height and rudder position were adjusted as needed for each participant’s comfort. Calibration to adjust the interpupillary distance of the HMD and the participant’s head position in the virtual environment then took place. All participants then received 3-min of free-play of simulated flying in the Sprint Chair to familiarize themselves with the virtual environment and its physical controls. Participants did not perform the OHB during this free-play period.

Each participant then performed the OHB 7 times in the experiment and graded by the Evaluator after each repetition. The Evaluator provided oral feedback to the participant before each trial as necessary, consistent with the evaluation criterion in Table 1. Examples of feedback by the Evaluator to the participant included ensuring that the landing gear and flaps were in the up position, and that torque was set at approximately 65% before each repetition. Each repetition was conducted from the same starting position at 3000 ft. altitude.

After the 7<sup>th</sup> OHB repetition, participants completed the post-experiment SSQ and participants in the Expert condition completed the SiFi scale in addition to the SSQ. The study took 45 min to complete. The experimental procedure is summarized in Figure 5 below.

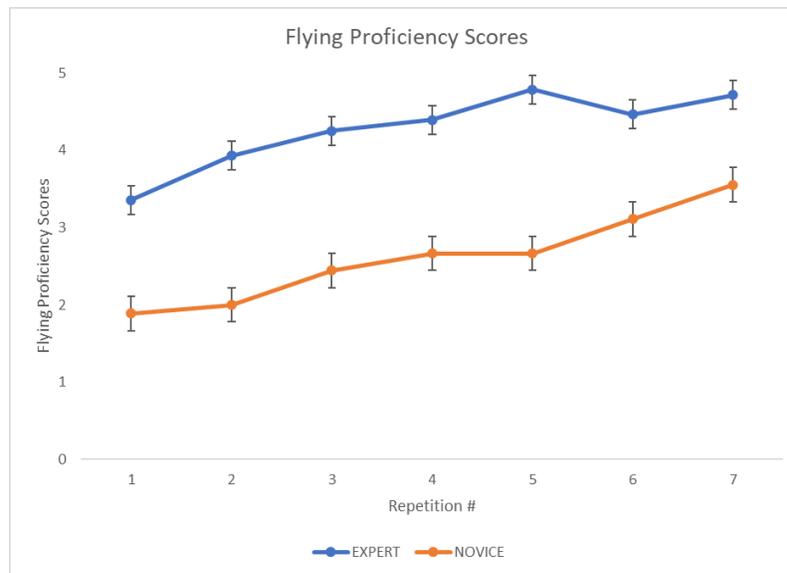


**Figure 5. Summary of experimental procedure.**

## RESULTS

### Expert v Novice Flight Proficiency Scores

All analyses were performed with a 0.05 two-tail  $\alpha$  cut-off in SPSS 28. A mixed factors ANOVA with Repetitions (1 through 7) as the within-subjects factor and Condition (NOVICE v EXPERT) as the between-subjects factor was performed. FPS for both conditions are displayed in Figure 1. The mixed factors ANOVA revealed a significant interaction effect between Condition and Repetition  $F(6, 126) = 2.55, p = .02, \eta^2 = .11$ . According to Cohen (1988), this represents a medium effect size for Scores between the NOVICE and EXPERT conditions for each Repetition (all  $ps < .001$ ). Data are presented in Figure 6.



**Figure 6. Flight Proficiency Scores over 7 repetitions for Expert (blue) and Novice (orange) pilots.**

### Expert and Novice Flight Proficiency Score Variability Over Time

As a result of this significant interaction effect, post-hoc pairwise comparisons with Bonferroni correction were performed on FPS over Repetitions. In the NOVICE condition, there were significant differences in FPS between repetitions 1 v 4-7; 2 v 4-7; 3 v 6-7; 4 v 7 and 5 v 7. In the Expert condition, there were significant differences in FPS between repetitions 1 v 2-7; 2 v 4-7; 3 v 5 and 4 v 5.

### Cybersickness Scores

A paired-samples t-test compared pre-post SSQ TS scores across all participants and indicated no significant difference,  $t(22) = -1.43, p > .05$ . Scores observed herein fall below the threshold of a ‘problem simulator’ and regarded as negligible cybersickness (Kennedy et al., 2003; Stanney, Graeber, & Kennedy, 2006). Data are displayed in Figure 7 below.



**Figure 7. SSQ scores before and after experiment.**

### SiFi to FPS Correlation

Spearman correlations were performed between average SiFi Scores and average FPS scores for each participant in the Expert Condition to understand the relationship between simulator fidelity and Expert pilot performance. Each of the 5 dimensions of the SiFi scale (e.g. auditory, visual, tactile, behavioral, and user interaction) was then correlated to the FPS scores, totalling 6 correlational analyses.

Findings indicated a significant negative correlation between FPS scores and Average SiFi scores,  $\rho(12) = -.547$ ,  $p = .043$  and User Interface scores,  $\rho(12) = -.062$ ,  $p = .017$  but not others.

### Expert Live Flight Hours and FPS Correlation

FPS were correlated with live flight hours of the aircraft being simulated in the Sprint Chair (i.e. the Harvard II) to determine if there was a relationship between live flight hours and performance in the Sprint Chair. Results indicated there was no significant correlation between number of flight hours and FPS in the Sprint Chair,  $\rho(12) = .348$ ,  $p > .05$ .

### SiFi to Live Flight Hours Correlation

Average SiFi scores to live flight hours were correlated to determine if there was a relationship between time spent/familiarity with the actual aircraft and perceived simulator fidelity. Results indicated there was no significant correlation between these,  $\rho(12) = -.201$ ,  $p > .05$ .

All Expert pilots had flown the Harvard II aircraft within days of this experiment, presumably negating any effect of skill fade from not having recently flown the real aircraft. Only a handful of participants were familiar with VR and used it on a regular basis, making it statistically unfeasible to run analyses comparing VR familiarity with relevant metrics such as SSQ scores and FPS collected in this experiment.

## DISCUSSION

The objective of the present study was to investigate the TE and validation of a VR HMD-based trainer for pilot training. This was done using an RToT design with Novice and Expert pilots performing the OHB. In RToT research designs, it is hypothesized that expert performance acts as validation of the system. That is to say, Experts performing at a high level in the training intervention supports its validation, whereas Expert poor performance indicates invalidation of the training intervention (AGARD, 1980; Korteling, Oprins, & Kallen, 2013). In the current experiment, FPS between Experts and Novices were compared. Findings indicated a significant interaction between Novice and Expert Scores on the OHB. Specifically, Expert scores were significantly greater than Novice FPS, consistent with the main hypothesis for the RToT design (AGARD, 1980; Korteling, Oprins, & Kallen, 2013). However, findings indicated that Experts performed significantly better in the Sprint Chair over time, likely reflecting a learning curve from the aircraft represented in the VR HMD-based simulator.

The second objective of this experiment was to determine if VR HMD-based simulators could be a useful training adjunct for rehearsing certain flight maneuvers. We hypothesized that if Novice pilots scores increased significantly throughout the experimental session, this would indicate evidence supporting VR-HMD based simulation for student pilots rehearsing the OHB when other training is not available. Data supported this hypothesis as students undergoing flight training that had never previously performed the OHB in the Novice condition showed significant performance improvements indicated by FPS increases throughout the experiment.

Cybersickness is known to be an important deterrent to the use of xR HMDs for military training and operations (Kaufeld et al., 2022; Kirollos & Jarmasz, 2021; Kirollos & Merchant, 2023; Proietti et al., 2021). Small increases in cybersickness were observed from the start to end of the experiment but these were not significant and negligible according to SSQ TS score categorization methods outlined by Kennedy et al. (2003) and Stanney, Graeber and Kennedy (2006).

The SiFi scale (Bush, Lamb, & Braun, 2022) was used to assess the relationship between fidelity of the Sprint Chair by Expert pilots relative to the Harvard II aircraft it represents. This was done by comparing SiFi scale ratings to FPS scores. Findings indicated that Average SiFi scores of all 5 dimensions combined correlated negatively and significantly with FPS. The User Interface dimension was the only dimension that also had a significant negative correlation to FPS. These findings likely indicate that increased pilot proficiency in the real aircraft results in less tolerance of differences in the simulator relative to the real aircraft. Live flight hours and FPS scores were also correlated in Expert pilots to determine the relationship between flying expertise and performance in the Sprint Chair, but though this correlation was positive, this was not significant.

### **Strengths, Limitations and Future Research**

A strength of this study was that it employed an RToT design that allows validation of a simulator without over-reliance on the training intervention, mitigating risks that an experimental training intervention can pose as in many traditional ToT designs. An additional strength of this study is that it had strictly defined definitions for Experts and Novices to best compare the difference in skills in the flying task in a VR HMD-based flight simulator.

However, there were also some important limitations that must be acknowledged when interpreting results from this experiment. First, data collection for this study is ongoing and preliminary findings are presented herein. Another limitation of this study is that only the OHB was performed. Similar studies should expand on the present study by evaluating and assessing the validity for TE of other flight tasks in VR HMD-based simulators. An additional limitation of this study is that clear daytime conditions were exclusively used in the flight scenario. Future research should attempt to replicate our findings in varying environmental conditions such as night and poor weather conditions.

An additional limitation in this experiment was that the PCL for the Warthog aircraft was used because a functioning Harvard II throttle was not available. This undoubtedly reduced simulator fidelity for Experts but was a constant error throughout the experiment. It is reasonable to assume that barring hardware issues, and reliable throttle function, the Harvard II replica throttle use in the Sprint Chair would have resulted in better simulator fidelity, learning transfer and possibly higher FPS scores than those observed herein.

### **CONCLUSIONS**

In this study, we present preliminary findings of our evaluation of a VR HMD-based simulator on the OHB in Novice and Expert pilots. Based on preliminary findings in this study we conclude that 1) VR HMD-based simulators, in this instance the Sprint Chair, may be a valid training adjunct based on expert pilot performance that 2) student pilots may be able to safely rehearse the OHB in the Sprint Chair. 3) Cybersickness appears to be negligible for durations of OHB training of approximately 45 min in VR. Future research by our group will present final data for this experiment and examine how the Novice Student cohort FPS scores change compared to archival data of students that did not use VR HMD-based simulation for rehearsing the OHB during live flight. This comparison can provide additional validation of VR HMD-based simulation for pilot training.

### **ACKNOWLEDGEMENTS**

The authors would like to thank the 2 CFFTS, Moose Jaw aviation community including all study volunteers, technical contractor staff, and Wing Command for their hospitality and support during data collection. In particular, the authors would like to thank Capt. Devin Mckenna, David Kanz and Brent Peardon for their critical assistance throughout data collection campaigns.

## REFERENCES

- AGARD. (1980). *Fidelity of simulation for pilot training: Annex A - Assessment of Training Effectiveness*.
- Allsop, J., Olonilua, M., & Yeo, S.-H. (2021). *Examining Quas-Transfer of Training From A High Specification Virtual Reality Training Device to a High Fidelity Fixed-Base, Dome Flight Training Device - Interim Report*.
- Bush, D., Lamb, C., & Braun, A. (2022). Interrater Reliability of the Simulation Fidelity (SiFi) Scale. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,
- Cain, B., Magee, L., & Kersten, C. (2012). Validation of virtual environments incorporating virtual operators for procedural learning.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* 2nd ed.(Hillsdale, NJ: L. Erlbaum Associates).
- Federal Aviation Administration. (2021). Air Traffic Control - Terminal, Arrival Procedures and Separation. In [faa.gov/air\\_traffic/publications/atpubs/atc\\_html/chap3\\_section\\_10.html](https://www.faa.gov/air_traffic/publications/atpubs/atc_html/chap3_section_10.html).
- Kaufeld, M., Mundt, M., Forst, S., & Hecht, H. (2022). Optical see-through augmented reality can induce severe motion sickness. *Displays*, 74, 102283.
- Kennedy, R. S., Drexler, J., Compton, D., Stanney, K., Lanham, D., & Harm, D. (2003). Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: similarities and differences. In *Virtual and Adaptive Environments*. CRC Press.
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3), 203-220.
- Kirollos, R., & Jarmasz, J. (2021). *Safety considerations for the land vehicle crew training system (LVCTS): An analysis of the literature*. (R2021-101). Defence Research and Development Canada
- Kirollos, R., & Merchant, W. (2023). Comparing cybersickness in virtual reality and mixed reality head-mounted displays. *Frontiers in Virtual Reality*, 4, 18.
- Korteling, H. J. E., Oprins, E. A. P. B., & Kallen, V. L. (2013). *Measurement of Effectiveness for Training Simulations* (RTO-SAS-095).
- McCoy-Fisher, C., Mishler, A., Bush, D., Severe-Valsaint, G., Natali, M., & Riner, B. (2019). *Student Naval Aviation Extended Reality Device Capability Evaluation*.
- Mishler, A., Severe-Valsaint, G., Natali, M., Seech, T., McCoy-Fisher, C., Cooper, T., & Astwood, R. (2022). *Project Avenger Training Effectiveness Evaluation* (ORL006).
- Munn, J., Jarmasz, J., & Deluce, D. (2023). *Comparison of Visualization technologies to Support Royal Canadian Air Force (RCAF) Training Modernization* IITSEC 2023,
- Proietti, P., Allsop, J., Bloch, M., Bos, J. E., Burov, A., Clement, D., French, J., Kirollos, R., Lawoson, B., Leoncini, P., Liarokapis, F., Olonilua, M., Sjolund, P., Rigato, P., Rodabaugh, T., & Secci, S. (2021). Guidelines for Mitigating Cybersickness in Virtual Reality Systems [Government Report]. *NATO STO*, 166.
- Royal Canadian Air Force. (2014). *Integrated Training Plan – Phase 2 Harvard*. Ottawa, Canada
- Royal Canadian Air Force. (2020). CT-156 Harvard II Aircraft [Photograph]. In: Department of National Defence Canada.
- Severe-Valsaint, G., Mishler, A. D., Natali, M. W., Astwood Jr, R. S., Seech, T. R., & McCoy-Fisher, C. E. (2022). *Training Effectiveness Evaluation of an Adaptive Virtual Instructor for Naval Aviation Training*.
- Stanney, K., Graeber, D., & Kennedy, R. S. (2006). Virtual environment usage protocols. In *Handbook of standards and guidelines in ergonomics and human factors*. Lawrence Erlbaum Associates.