REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
31-05-2022	Technical Report	1 October 2019 - 30 September 2021

4. TITLE AND SUBTITLE
Training Effectiveness Evaluation of an Adaptive Virtual Instructor for Naval Aviation Training

5a. CONTRACT NUMBER
N68335-19-C-0510

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Gabriella Severe-Valsaint, MS
Ada Mishler, PhD
LT Michael Natali, PhD
Randolph Astwood, PhD
LT Todd Seech, PhD
Cecily McCoy-Fisher, PhD

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Air Warfare Center Training Systems Division
Chief of Naval Air Training

8. PERFORMING ORGANIZATION REPORT NUMBER
NAWCTSD Public Release 22-ORLO44

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
CDR Brent A. Olde, MSC, USN, Ph.D.
Lead, Air Warfare Training Development IPT
Strategic Planning and Science & Technology, PMA-205
47123 Buse Road
Building 2272, Room 345

10. SPONSOR/MONITOR'S ACRONYM(S)
PMA-205

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
NAWCTSD Public Release 22-ORL044 Distribution Statement A – Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The Navy investigated VR training capabilities within their aviation training and results showcased significant promise for using VR devices, but also highlighted limitations reducing its potential benefit to training (McCoy-Fisher et al., 2019). Two critical issues identified were the need for guidance on the what, when, and how to train with VR and feedback on performance. As a potential solution, both the Air Force and the Navy started work in artificial intelligence (AI) instruction applied to flight training. This study examines an experimental version of an adaptive virtual instructor, the Virtual Instructor Pilot Exercise Referee (VIPER®) from Discovery Machine Inc. (DMI), a first step towards an AI instructor Pilot capability.

For the Navy to better understand the training impact VR and introduction of an AI-style tutor may have, this study evaluates whether there were performance improvements in student naval aviator (SNA) flight events across different training conditions utilizing the Navy's Immersive Training Devices (ITDs). There were four conditions in this study:

15. SUBJECT TERMS
virtual reality, adaptive virtual instructor , virtual instruction, artificial intelligence, intelligent tutoring system, primary training, evaluation, VIPER

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*

NAVAL AIR WARFARE CENTER
TRAINING SYSTEMS DIVISION

ORLANDO, FL 32826-3275

NAWCTSD Public Release 22-ORL044

26 May 2022

Experimental and Applied Human Performance Research & Development
Technical Report

Training Effectiveness Evaluation of an Adaptive Virtual Instructor for Naval Aviation Training

by

Gabriella Severe-Valsaint, MS
Ada Mishler, PhD
LCDR Michael Natali, PhD
Randolph Astwood Jr., PhD
LT Todd Seech, PhD
Cecily McCoy-Fisher, PhD

Prepared for:
Naval Air Systems Command (NAVAIR)
PMA-205 Naval Aviation Training Systems and Ranges
Patuxent River, MD 20670

DR. JAMES PHARMER
Chief Scientist, Research, Development, Test and Evaluation Department (GT5E)

WILLIAM ZELLER
Director, Research & Technology Programs

Disclosure

Contents

1. Acknowledgments

For acquiring the funding, coordinating across CNATRA, scheduling and supporting data collection, and providing needed expertise, we thank the following individuals:

Naval Aviation Training Systems and Ranges
(PMA-205)
CAPT Lisa Sullivan
CDR Brent Olde
CDR Chris Foster
LT Joseph Mercado (retired)
Kirklin Rhodes
Jeffrey Llewellyn
Gary Moser

Chief of Naval Air Training
(CNATRA)
RDML Robert Westendorff
CAPT Kevin Delano
CAPT Steven Hnatt (retired)
COL Austin Burrill (retired)
CDR Kerry Bistline
CDR Joshua Calhoun
LCDR Anthony Matheus
LCDR Kwame Anyika
Thomas Sheppard
Jessica Richards
Thomas Ford
Justin Wallace
Gloria Elizondo
Richard Garms

Naval Air Warfare Center Training Systems Division
(NAWCTSD)
CAPT Daniel Covelli
CAPT Timothy James
CAPT Timothy Hill (retired)
Brian Hicks
Gregory Ouellette
Tyson Griffin
Dr. James Pharmer
Dr. Heather Priest-Walker
Beth Atkinson
Brian Hicks

2. Executive Summary

2.1 Problem and Objective

To address the on-going pilot shortage, the Department of Defense (DoD) services are exploring the capabilities of virtual reality (VR) technology to improve the efficiency and effectiveness of aviation training. Although the Air Force found positive results and feedback from leveraging VR in their experimental Pilot Training Next (PTN) Program, there has been limited published results on the impact to flight performance. Parallel to this, the Navy investigated VR training capabilities within their aviation training and results showcased significant promise for using VR devices, but also highlighted limitations reducing its potential benefit to training (McCoy-Fisher et al., 2019). Two critical issues were identified in the Navy's investigation; first, the need for guidance on the what, when, and how to train with VR and second, the need for feedback on performance when practicing in VR. As a potential solution to address the concerns above, both the Air Force and the Navy started work in artificial intelligence (AI) instruction and feedback for application within VR flight trainers. This study examines an experimental version of a virtual adaptive instructor, the Virtual Instructor Pilot Exercise Referee (VIPER®) from Discovery Machine Inc. (DMI), a first step towards an AI instructor Pilot capability.

For the Navy to better understand the training impact VR and introduction of an AI-style tutor may have, this study evaluates whether there were performance improvements in student naval aviator (SNA) flight events across different training conditions utilizing the Navy's Immersive Training Devices (ITDs). There were four conditions in this study:

1. *Archival*: no access to ITDs,
2. *Free VR*: free-play access to ITDs without guidance or VIPER®,
3. *Assigned VR*: required practice scenarios in the ITDs without VIPER®, and
4. *VIPER®*: required practice scenarios in the ITDs with VIPER®.

Results provide quantitative data on the effectiveness between groups demonstrating VR benefit to training and qualitative feedback on the utility and usability of a virtual adaptive instructor, in this case VIPER®.

2.2 Method, Assumptions, and Procedures

Participation in the study required students to practice for a minimum of 18 hours on the devices within one of two experimental conditions, Assigned VR and VIPER®, while archival data were used for the other two conditions. SNAs were provided an overview of the study requirements for participation and instructions on how to use the VR devices. SNAs were instructed to engage in practice on the devices during their free time, not to interfere with their training schedule. Students had to log their practice to monitor hours and issues encountered during their sessions. At the completion of data collection, performance data on flight events were acquired for all four groups and usability and utility feedback were captured from both instructor pilots (IP) and students. Performance and feedback data were analyzed for trends and recommendations.

2.3 Results

A total of 292 SNAs were recruited at the beginning of their Primary Training for the two experimental groups at Training Wing Four (TW4), NAS Corpus Christi, Texas. Unfortunately, due to dropout rates, the team received completed datasets from only 64 SNAs for the Assigned VR condition and 52 SNAs in the VIPER® condition (116 SNAs total). Archival data (i.e., Archival and Free VR) totaled 4,179 SNAs for comparison. Additionally, feedback data on VIPER® were received from the two IPs involved in the program and 15 of the SNAs who participated in the study.

2.3.1 Quantitative Results

To compare performance across conditions, grades (event raw score or ERS) and extra training events (event modifier code) were used to determine any differences between the four groups in the Contact and Instruments Phase of Primary Training. Mann-Whitney *U* tests comparing the four groups were conducted to identify grade and modifier code differences.

Event Raw Score

Comparisons of individual event grades showed statistically significant performance increases[1] between each consecutive training intervention and within every event examined in the two phases. Specifically,

- The Free VR condition had a significant overall increase in ERS compared to the Archival group in 76% of events (84% of Contact events and 64% of Instruments events).
- The Assigned VR condition had a significant overall increase in ERS compared to the Free VR group in 42% of events (47% of Contact and 36% of Instruments).
- The VIPER® condition had a significant increase in ERS compared to the Assigned VR group in only one Contact event (C4304; 5% of Contact, 0% of Instruments, and 3% of overall events).

At an aggregate level (i.e., average ERS across Contacts and Instruments) each successive level had a significant increase in scores with up to approximately half a standard deviation higher performance. Equating these effect sizes to Navy Standard Score (NSS) metrics (a 20 to 80 scale with standard deviation of 10), the increases in performance scores ranged from 3.3 to 6 NSS points depending on the condition.

Event Modifier Codes

Findings from the comparison of the modifier codes were not as straightforward as the comparisons of grades, but did follow a similar pattern. VIPER® and Assigned VR had lower occurrences of events with modifier codes than the two archival conditions, where the Archival group had the most modifier codes present. Unfortunately, statistical comparisons could not be performed between all conditions due to groups with zero modifiers; limiting study results.

Additionally, a Fisher's Exact Test was conducted to compare failed events (i.e., unsats) for all four groups. Out of the modifier codes analyzed, unsats are the most tied to SNA performance and therefore warranted an independent analysis.

---

[1] Note: To address familywise error rates related to multiple pairwise comparisons, alpha was set to .01 for the purposes of the statements made in this section. Close attention should be given to Appendix 1, in which significance levels, effects sizes, and inferred power achieved together can provide more precise results.

Results showed the Archival group received more unsats than the conditions with VR exposure in the Contact Phase and Overall, but only received significantly higher unsats than the Free VR group in the Instruments Phase.

2.3.2 Qualitative Results

SNA Feedback

SNAs were asked to provide feedback on their experience with the virtual instructor. The team received responses from 15 out of 52 students in the VIPER® group. Although there was a low response rate and many comments focused on initial program crashes (later resolved), 93% of students who responded to the questionnaire expressed potential for the virtual instructor to support skills across every chapter of the syllabus. Based on those 15 respondents:

- 20% of students stated VIPER® could help learn course rules,
- 53% stated the system helped provide sight pictures,
- 80% expressed it better prepared them for upcoming events, and
- 64% to 92% of SNAs, depending on the specific question, rated the maneuvers practiced as slightly effective or higher.

These offer support to the benefit of training with the system. SNAs also reported the system was easy to use and seemed to provide both timely and accurate instructions and feedback. Finally, though there were some issues with VIPER® understanding SNA auditory commands, the visual and auditory delivery of instruction and feedback were clear and easy to comprehend.

IP Feedback

Responses from the IPs were promising as well, indicating VIPER® demonstrated value as an early training tool. Specifically, IPs expressed that the system could provide benefit for developing scan patterns, engaging in self-study prior to flying, and is capable of honing skills for some maneuvers. However, IPs also commented on grading inflexibility and voice recognition limitations associated with VIPER®. IPs indicated that for certain maneuvers (e.g., landing pattern), grading and feedback were less useful because the maneuvers can be completed successfully in multiple ways, requiring a more flexible grading rubric. Additionally, despite recognizing its benefits, the majority of IPs did not view the system positively. This perspective may have been formed early on when VIPER® was experiencing various technical issues and continued even after improvements to the system were made. Unfortunately, IPs indicated that when VIPER® was discussed, system frustrations were mentioned which led to a negative reputation of VIPER® that affected both IP engagement and SNA participation.

2.4 Recommendations

Overall, virtual adaptive instructor programs such as VIPER® have the potential to improve pilot performance and this evaluation was a first step towards providing objective data for incorporating AI instruction into flight training. To ensure maximum benefit of VIPER® or similar programs, responses from IPs and SNAs were summarized to provide recommendations for future development and integration within aviation training:

- Auditory: improve voice command recognition and response accuracy; or provide an alternative to auditory inputs that do not require the navigation of drop downs with a mouse.

- Maneuver Development: expand the maneuvers available to practice and introduce more flexibility in grading non-standard but acceptable ways of completing a maneuver. Ensure early IP participation for accurate modeling of maneuvers, feedback type, and feedback delivery.

- Feedback Delivery: provide more theoretical information about each maneuver (e.g., when and why a maneuver should be completed), show feedback for partially completed maneuvers, and provide after-action feedback in a better format (e.g., show percentage of the maneuver correctly completed).

- Implementation: include VIPER® on low-cost simulators as a form of pre-Primary self-practice for SNAs who have completed or during Naval Introductory Flight Evaluation (NIFE), provide both IPs and SNAs an overview of the system and its capabilities, and invest in a more accurate flight model to increase VIPER®'s utility (e.g., aerobatic maneuvers).

3. Introduction

3.1 Problem

In recent years, the Department of the Navy has placed an increased focus on utilizing emerging simulation technology to help supplement current aviation training. More specifically, the Navy is exploring ways to increase training efficiency and effectiveness to address the Fleet pilot shortage (United States Government Accountability Office, 2018). In response to this need, from 2018-2019 the Naval Air Warfare Center Training Systems Division (NAWCTSD), Naval Aerospace Medical Institute, Chief of Naval Air Training (CNATRA), and Naval Aviation Training Systems and Ranges (PMA-205) collaborated on a study to examine the potential benefits and impacts of supplemental Virtual Reality (VR) practice on student pilot performance in the aircraft (McCoy-Fisher, Mishler, Bush, Severe-Valsaint, Riner, & Natali, 2019). Although results identified advantages regarding these VR trainers, minimal documentation of student practice time and little guidance or structure on how or what to practice prevented robust conclusions on the potential for performance improvements within training.

As a follow-on study and further development for the devices, PMA-205 and CNATRA collaborated to leverage work conducted with the Air Force in artificial intelligence (AI) instruction for their experimental Pilot Training Next (PTN) Program. Via the Small Business Innovation Research (SBIR) program, the Air Force funded Discovery Machine, Inc. (DMI) to develop the Virtual Instructor Pilot Exercise Referee (VIPER®) to support flight maneuver practice in their virtual T-6A devices. Due to initial positive feedback on VIPER's capability and potential to increase learning gains, the Navy utilized the SBIR program to fund DMI via a Phase II SBIR to develop a T-6B version of VIPER® to support Naval Aviation Training Next (NATN) and Primary flight training. The VIPER® program was incorporated into CNATRA's T-6B Immersive Training Devices (ITDs; VR trainers made of commercial off-the-shelf [COTS] components on desktop computers) at Naval Air Station (NAS) Corpus Christi, TX. The Multidisciplinary Extended Reality (MXR) research team at NAWCTSD was funded to evaluate the impact of VIPER® on student naval aviator (SNA) performance.

3.2 Objectives

The purpose of this evaluation was to assess if the additional capabilities presented by an AI or adaptable virtual instructor, in this case VIPER®[2], would improve SNA performance beyond traditional training with no ITD access or training with ITD access. Specifically, the goal of this evaluation was to determine if there were differences in performance across four SNA conditions:

- *Archival*: no access to ITDs;
- *Free VR*: free-play access to the ITDs without VIPER®;
- *Assigned VR*: required practice scenarios in the ITDs without VIPER®;
- *VIPER®*: required practice scenarios in the ITDs with VIPER®.

Although some may consider VIPER® too early in development to be considered a fully AI program, it is an initial step towards AI instruction for flight training and a test of its potential utility. Findings from this evaluation will provide a better understanding of VR and AI technologies' ability to support Naval aviation training and indicate any additional development needed for virtual instructors, such as VIPER®, and future AI instructional programs to be most beneficial for Primary Training.

3.3 Background

3.3.1 T-6B ITD

VR technology, employed as part of an ITD, is being explored and evaluated as a new way to provide aviators with supplemental training at a lower cost, where a single high-end[3] ITD can cost as low as $50k[4] while traditional operational flight trainer simulators cost in the hundreds of thousands to millions of dollars. In 2018, CNATRA acquired its first version of a T-6B

---

[2] *Note*: Though the basis of the research and development work is in and for an AI instructor, this report will refer to VIPER® as an adaptive virtual instructor to avoid any technical disagreements or misunderstanding of official definitions or distinctions of AI instruction.

[3] Low-end ITD trainers can cost as little as $10-20k.

[4] The $50k estimate only includes the hardware and software; additional costs are incurred when accounting for maintenance and sustainment of devices.

ITD (based on the Air Force’s T-6A version), representative of the Beechcraft T-6B Texan II aircraft used for Navy Primary Training. Each ITD consisted of a desktop computer, monitor, COTS components (gaming seat, VR headset, hands on throttle and stick [HOTAS], and rudder pedals), and a virtual environment based on commercially-available software.

The Navy’s initial introduction of the ITDs focused on device exposure and “free play” for students, providing greater opportunities for immersive study, practice, and repetitions of skills that are often described as the “reps and sets” necessary for skill development. The what, when, or how to practice was at each student’s discretion unless an instructor voluntarily offered any recommendations or guidance on use of the device. Additionally, performance monitoring and feedback relied on students’ own knowledge, reflection, and recognition of their current state of performance. In other words, SNAs had to recognize their own mistakes and understand how to correct them. Overall, the ITDs provided SNAs an immersive platform to practice and develop skillsets with the ability to see in “real time” the effects of their actions instead of the traditional desk study or “chair flying” with paper printouts.

Results from the McCoy-Fisher et al. (2019) study of the initial introduction of the ITDs provided promising support on the benefits of VR device use but also identified necessary upgrades to optimize the new technology’s impact on training performance. In particular, for Primary flight training, the study found VR to be most useful for building a sight picture of upcoming events and practicing skills relevant to the Contact phase of the syllabus – support that the immersive environment was beneficial to training. Grade data were not available for T-6B ITD users, but examination of similar T-45C Goshawk ITDs found increased flight performance in some phases of the Advanced Strike syllabus, indicating that using ITDs may enhance performance in live flight.

Findings also identified several upgrades to implement in order to better leverage the technology and provide greater training benefit. Most updates focused on hardware and software upgrades to improve flight characteristics, visual fidelity, and control feel to increase the accuracy and realism of the device (i.e., looks, feels, and acts like the actual aircraft). Additionally, two other major improvements were identified:

1. The need for structured guidance on how, when, and what to do with the device. Practice in the device should be deliberate with a certain focus or objective(s);

2. The need for timely feedback on performance when utilizing the device, either by a human or computer-based instructor, to ensure good habits are learned and poor performance is identified and corrected.

These coincide with extensive research on developing expertise, specifically on the benefits of deliberate practice and feedback. It also demonstrates that technology alone is insufficient to improve training; it needs appropriate integration to maximize its benefit.

3.3.2 Developing Expertise

Defined as the acquisition of superior, reproducible performance in a particular domain (Ericsson & Charness, 1994), achieving expertise across a range of skillsets is the ultimate goal of Naval Aviation Training: "to safely train the world's finest combat quality aviation professionals" (CNATRA, 2022). To develop expertise requires significant time and effort, but this alone is insufficient and how time and effort are applied matters (Ericsson & Charness, 1994). Specifically, research shows there are four conditions that facilitate expertise development:

- Well-defined goals;
- Motivation to improve;
- Provided with feedback;
- Provided ample opportunities to practice.

SNAs are generally motivated to complete flight training and the ITDs improve available practice opportunities. Where use of the ITDs can be improved are the other two conditions: well-defined goals in relation to practice (i.e., "deliberate practice") and some type of feedback mechanism for the student.

Deliberate practice

Deliberate practice is more than free play or mindless repetition of a task. It is attentive, effortful practice aimed at improving performance on specific skills, and it requires: learning successful ways of completing the task; feedback on current state in relation to a set goal or standard, progress made towards the goal or standard, and strategies for corrections and improvements; and high repetition specifically with the intention of addressing or incorporating feedback to refine performance. Ericsson, Krampe, & Tesch-Römer (1993) found a direct relationship between the amount of deliberate practice people engage in and their level of performance, showing it to be a primary determinant of expert status in the practiced domain. Not only does the amount of deliberate practice accumulated affect current performance, but the amount of time currently spent in deliberate practice also distinguishes between relatively good and poor performance among experts (e.g., continually refining skills via deliberate practice vice only practicing already mastered skills). Additionally, to continually develop a skill to reach expert levels, deliberate practice should take the form of individual tasks slightly more difficult than the trainee's current ability level, and as performance improves on the individual tasks, they are combined into more complex scenarios.

Integral to deliberate practice is the need for the trainee to know what and how to practice, a need generally served by the presence of an instructor, and research has demonstrated instruction can help poor performers catch up with better performers (Ericsson, 2008; Ericsson & Charness, 1994; Ericsson, Krampe, & Tesch-Römer, 1993). The feedback from an instructor provides the trainee with information on what and how to improve, guiding goal setting and practice strategies. Getting feedback on performance is not limited to instructors but can be accomplished via comparing one's own performance with experts' performance either via self-monitoring and reflection, objective measures, or from someone knowledgeable on the domain (Ericsson, 2008).

Feedback

As the work above illustrates, practice alone is insufficient to improve and sustain performance, to maximize training benefit requires feedback. Feedback serves a specific purpose: identify discrepancies between current state and desired end state (i.e., the goal or standard) as well as provide potential avenues for reaching the desired end state to facilitate learning and skill development. However, there are a number of ways in which feedback can vary: 1) the content of the feedback (e.g., outcome, process, normative), 2) the feedback sign (e.g., positive or negative), 3) the modality of feedback delivery (e.g., orally or written), 4) the amount of feedback given, and 5) the timing (e.g., delayed or immediate). Each of these dimensions has an impact on the effectiveness of feedback (Kluger & DeNisi, 1996; Kozlowski, Bell, & Mullins, 2000). Research supports feedback’s importance to learning: it reduces learners’ cognitive load, uncertainty in performance, and mistakes and errors; and potentially helps improve motivation (Billings, 2012). Additionally, research has examined the impact of feedback on performance finding that, in order to be effective and have a positive effect, feedback should follow certain guidelines (Billings, 2012; Ilgen, Fisher, & Taylor 1979; Kluger & DeNisi, 1996):

- *Clear and specific*: the feedback needs to be accurately perceived and understood for the receiver to take the appropriate actions.

- *Non-attributional*: the feedback should focus on the task, process, or behavior, not the person in order to keep attention focused on actionable changes to reach the goal or standard.

- *Credible source*: the receiver needs to trust the source of feedback is providing accurate and useful information either via sufficient knowledge, experience, and/or having observed the event providing feedback on.

- *Timeliness*: feedback should be delivered in a timely manner relevant to the complexity of the task(s) (simple vs. complex); the characteristics of the individual (novice vs. expert); and structure of the event (delivered during or after the event).

- *Individual needs*: the feedback should be suited to the type of task (simple vs. complex) and characteristics of the learner (novice vs. expert).

Based on these complexities, it is not surprising that inappropriately applied feedback can cause decrements in performance (Kluger & DeNisi, 1996). For example, understanding the aptitude of the learner can shift the appropriate method from a bottom-up approach for novices, whereby detailed feedback is initially provided on subcomponents of the task, and shifts over time into general feedback regarding the whole task, to top-down feedback for more advanced learners, beginning with an overview of the entire task and moving into detailed feedback on task subcomponents (Billings, 2012). As discussed below, the feedback provided by VIPER® was developed to fit all these criteria.

Ericsson et al. (1993) also found that feedback can be motivating. While deliberate practice alone may not be enjoyable, seeing improvements in one's own performance can be enjoyable and motivate people to engage in deliberate practice. Similarly, Kluger and DeNisi (1996) theorized that feedbacks impact on performance is mediated by motivation. Thus, feedback is important for performance not only directly, by affecting understanding of what and how to perform, but also indirectly by motivating people to improve their performance further.

Demonstration

Another critical component to developing expertise is the use of demonstration (also known as observational learning or modeling) to help learners understand what expert performance or the correct method looks like. Generally speaking, demonstration is considered a "dynamic example of partial-or whole-task performance of the characteristics of a task…that illustrates (with video recording, modeling, or any visualization approach) the enactment of targeted knowledge, skills, or abilities" (Salas et al., 2009 p. 2). In other words, demonstration shows the individual what "right" looks like. With theoretical foundations based on social cognitive theory (Bandura, 1986), nearly every modern organization, including Naval aviation, utilizes demonstration to great success via behavioral modeling training (BMT) to develop trainee skills (Taylor, Russ-eft, & Chan, 2005). To be effective, BMT relies on five primary components (Decker & Nathan, 1985; Salas et al., 2009):

- A list of well-defined skills and/or facts to be learned;

- Utilizing models and visual aids to illustrate effective behavior and skills;
- Opportunities to practice newly demonstrated skills;
- Feedback on practice performance related to what was demonstrated;
- What was demonstrated and learned is reinforced in follow-on applications (e.g., other training exercises or real-world scenarios).

However, utilizing these components does not guarantee the demonstration or BMT will be successful. Like practice or feedback, demonstrations can have a negative impact when used incorrectly or when reinforcing incorrect actions. As Salas and colleagues (2009) note, "the effectiveness of demonstrations depends upon the interrelationships between features of the demonstration, the learner, and the larger training system" (p. 12). Careful consideration of the type of task (simple vs. complex), level or style of demonstration (partial vs. whole task; video vs. live; minimal vs. expert performance), and learner characteristics (novice vs. expert) is needed to develop effective demonstrations.

Research supports the benefits of demonstration, deliberate practice, and effective feedback for improving performance where each one relies on and enhances the other two ultimately facilitating the development of expertise. By integrating them into the Navy's use of the ITDs in aviation training, it is expected to provide better learning and performance gains for students. However, with limited human instructor resources, CNATRA, NAWCTSD, and PMA-205 are investigating intelligent tutoring systems (ITS), such as adaptive instruction programs like VIPER®, to capitalize on demonstration, deliberate practice, and performance feedback benefits without requiring a human presence.

3.3.3 Intelligent Tutoring Systems

An ITS is defined as a system that aims to provide customized instruction and/or feedback to a learner without human intervention (VanLehn, 2011). These systems typically leverage instructional strategies identified by research (e.g., deliberate practice and feedback) to determine which training interventions to incorporate based on the learning objectives, individual needs, and performance level of the learner. ITSs are commonly referred to as regulative loop systems where performance is monitored, compared to a "gold standard" or a level of performance to be reached, and instruction is adjusted to get the learner closer to that standard based on performance levels (VanLehn, 2016); for example, by adjusting task difficulty, feedback type, or feedback timing. Simply put, these regulative loops consist of up to four components:

1. Self-regulation – the learner must determine their performance deviation from the standard;

2. Mirroring – the system provides a playback of the learner's performance for comparison to a set standard (e.g., an expert model);

3. Formative assessment – the system monitors and compares learner's performance to a set standard showing any discrepancies;

4. Coaching – the system monitors and compares performance to the standard and generates advice to modify learner's performance towards achieving the standard.

In order for ITS system designers to build effective systems, understanding when and how to use the four components above is critical (Billings, 2012; VanLehn, 2016).

VIPER®

VIPER® is an adaptable virtual instructor that helps tie in the demonstration, deliberate practice, and timely feedback aspects for developing expertise as training interventions when using the ITDs to promote knowledge and skill retention in the aviation community. In support of these interventions, VIPER® provides an individualized approach to instruction intended to mimic human instructors on six main attributes, as listed in DMI's proposal to CNATRA:

1. Understand the many ways things should be done;

2. Monitor trainees over time;

3. Assess trainee performance in real-time;

4. Identify when to intervene;

5. Identify how to intervene and act upon it;

6. Conduct After-Action Review. (Discovery Machine, Inc., 2019, p. 3).

The system aligns these attributes by applying expert mental models derived from human IPs to instruct students on various flight skillsets (Discovery Machine, Inc., 2019). From these models, VIPER® allows students to interact with the system in three modes:

1. Demonstration mode: the maneuver is selected and the tutor walks through a video of the maneuver explaining how to perform it.

2. Practice mode: the maneuver is selected, practiced, and feedback on performance is provided by the tutor.

3. Performance mode: the maneuver is selected but performed unassisted, the system identifies it was attempted, and performance is assessed with the assessment provided upon completion of the session.

These three modes align with the crawl-walk-run method commonly used in training. The system tracks students' progress over time and adapts its speech-based and text-based feedback according to their proficiency level on maneuvers in previous sessions over time. The VIPER® system also provides a speech interface for students to interact with the system via commands and questions and a tablet interface to track performance, select premade scenarios, or build their own scenario. It is important to note not all features were fully developed or used during this evaluation. Specifically, the performance mode and a separate instructor-only interface were among those not utilized for this study.

These VIPER® capabilities leverage expertise research literature on demonstration, deliberate practice, and feedback in the following ways:

1. The presence of preset maneuvers and the use of demonstration mode allow SNAs to understand what they should be practicing in the ITD. Although a live

instructor would still be ideal, VIPER®'s high availability via the ITDs can provide SNAs structured guidance for their practice during non-scheduled hours or when IPs are otherwise occupied.

2. By explaining how to perform a maneuver, demonstration mode helps users understand performance standards, set clear goals for their performance, and sets the stage for accurate performance discrepancy judgments.

3. Allows for both individual maneuver practice and practice of scenarios composed of strings of maneuvers, providing opportunity for increased complexity for events based on individual learning level and performance to aid expertise development (Ericsson, 2008).

4. Provides timely feedback leveraging the guidelines above derived from the research literature (Billings, 2012; Ilgen et al., 1979; Kluger & DeNisi, 1996). That is, VIPER® gives feedback related to the task, in the form of specific components of the task that were not completed correctly, as well as how much over or under the ideal value they were. The use of over/under values provides specifics on the difference between current and desired state and allows SNAs to understand what they should change to meet ideal performance (thus meeting the need for feedback that enables the selection of the correct answer). The use of this detailed subtask feedback also serves the purpose of Primary Training well by providing novice pilots with the type of feedback best suited to their early training (Billings, 2012).

5. Summaries of SNA's previous performance on a given maneuver is provided before the start of the current attempt, which helps SNAs judge how their performance changed from the previous attempt (i.e., whether their corrective actions are working to improve performance).

6. Maneuvers are based on the input of expert pilots, which allows SNAs to compare their performance to expert performance (the goal or standard), in alignment with Ericsson's (2008) recommendation.

With these features developed, VIPER® should provide opportunities for demonstration and deliberate practice with effective feedback that aligns with the expertise and learning science to assist SNAs in improving their flight skills.

3.3.4 Hypotheses

To examine how VR practice and adaptable virtual instruction may benefit SNA flight performance in Primary Training, the research team compared four separate groups of students:

1. Archival: SNAs with no access to ITDs;
2. Free VR: SNAs with free access to ITDs without any structured guidance or VIPER®;
3. Assigned VR: SNAs assigned to complete specific practice scenarios in the ITDs without VIPER®;
4. VIPER®: SNAs assigned to complete specific practice scenarios in the ITDs with VIPER®.

Those with free access to the ITDs may have used them, but the low usage of ITDs reported in McCoy-Fisher et al. (2019) suggests mean ITD usage in this group likely did not exceed a few hours across multiple months of training. Thus, it was expected that a higher level of ITD usage, in a more structured format, would lead to greater training benefits for those required to use the ITDs. In turn, VIPER® usage was expected to have higher benefits than ITD usage alone due to the guidance and feedback provided by the virtual instructor. Therefore, it was hypothesized that SNAs with VIPER® practice would be the highest performing SNAs followed by the assigned VR group, then the free VR group, and finally the archival SNAs as the lowest performing.

4. Methods

4.1 Participants

All data for the evaluation were collected from personnel located at Training Wing Four (TW4), NAS Corpus Christi, Texas.

4.1.2 Student Naval Aviators

This evaluation included a total of 292 SNAs recruited at the beginning of Primary Training, as well as archival data from 4,179 SNAs. The study used a convenience sample based on training class schedules and system availability to avoid impacting active training production.

All participants were provided an introductory session on setup and use of the ITDs as well as description of the study. The recruited participants were assigned to one of two conditions based on the timing of cohort class start dates and the availability of the systems to support each condition: the first 158 SNAs were placed in the Assigned VR condition (practice in the ITDs without VIPER®), and the subsequent 134 SNAs were assigned to the VIPER® condition (practice in the ITDs with VIPER®). However, due to a significant dropout rate, final data received were 64 SNAs for the Assigned VR condition and 52 SNAs in the VIPER® condition. Data collected from these two groups included performance data from the Training Sierra Hotel Aviation Readiness Program (T-SHARP) grade tracking system, weekly VR participation logs, and responses to a VIPER® questionnaire from SNAs in the VIPER® condition.

The evaluation also included archival performance data from CNATRA’s T-SHARP grade tracking system that were split into two groups: Archival and Free VR. The Archival group contained 850 SNAs who completed Primary Training before October 2018, when the ITDs were delivered, and therefore had no ITD access. The Free VR group contained 3,329 SNAs who began Primary Training after October 2018, and therefore had access to the ITDs for practice from the start of their training, but had no requirement to use the ITDs or guidance on how to use beyond basic startup procedures. However, the archival dataset did not include data relevant to the research questions for all SNAs; therefore, 836 Archival SNAs and 3,014 Free VR SNAs were included in analyses.

Thus, the four groups of SNAs in this study have progressively incorporated more aspects from the expertise and learning science literature:

- Archival (no ITDs): traditional, baseline training;
- Free VR (ITDs available but not required): provides increased opportunities for SNA self-directed practice;

- Assigned VR (required to practice in ITDs): provides SNA deliberate practice but relies on self-monitoring for feedback;

- VIPER® (required to practice in ITDs with VIPER®): provides SNA with demonstration and deliberate practice with an ITS delivering performance feedback to facilitate skill development.

It is important to note that this study had a high attrition rate. For the Assigned VR group, dropout rates may have been attributed to SNAs having competing training priorities as well as SNAs not completing their hours during the data collection timeframe; data were used from those who completed their 18 hours of practice. By contrast, students' requests for withdrawals from the VIPER® condition were heavily influenced by initial software instability issues associated with the first Navy version of VIPER® and the resultant frustration associated with interacting with a system under development. Fortunately, the system instability was addressed prior to completing data collection, but it remained difficult to recruit and maintain VIPER® participation throughout the remainder of the study. Based on Informed Consent Documentation, 94 SNAs (59%) in the Assigned VR condition and 82 (61%) in the VIPER® condition either withdrew from or did not complete the study. Data presented in this report only include participants who completed study requirements.

4.1.3 Instructors, Stakeholders, and Leadership

The research team also collected feedback from instructors through a wrap up questionnaire towards the end of the study. Although stakeholders and leadership were invited to participate, out of the eight solicited for feedback, only the two IPs who were involved enough to be familiar with the VIPER® program, development, and evaluation responded. They provided feedback on VIPER®'s capabilities and limitations as well as providing recommendations for future development and integration into the syllabus.

4.2 Materials and Apparatus

4.2.1 Materials

To prepare for the study, the research team in collaboration with IPs developed and distributed a T-6B Curriculum Breakdown Survey; a T-6B VIPER® Maneuver Feedback Questionnaire; and Participant Binders containing study materials, a syllabus outlining the practice scenarios, and the T-6B VIPER® SNA Questionnaire. In addition, IPs developed an Introduction Session for Assigned VR and VIPER® participants.

T-6B Curriculum Breakdown Survey

This survey was developed to capture initial feedback from IPs on what phases of training VIPER®'s capabilities would best support (see Appendix 4). The survey is sectioned off into the five phases of the Primary syllabus (Ground, Contacts, Instrument, Navigation, and Formation). IPs were asked whether or not VIPER® could support each training block within the five phases with response options of "yes," "no," and "maybe" and were also asked to explain their responses. These data were used to inform maneuver development within the VIPER® system and scenario development for the study most appropriate for the curriculum.

T-6B VIPER® Maneuver Feedback Questionnaire

The IPs were asked to provide feedback on initial maneuvers developed by DMI and verified for accuracy by CNATRA (see Appendix 5). IPs provided feedback on 33 maneuvers by first flying those maneuvers and then answering questions about VIPER® accuracy and effectiveness. For example, how accurate was VIPER® at: monitoring the aircraft, providing instruction prior to maneuver, and providing feedback upon completion of maneuver. These questions were rated on a 6-point Likert scale from "not accurate at all" to "extremely accurate." The survey also asked about effectiveness of student instruction on a 6-point Likert scale ranging from "not effective at all" to "extremely effective." The questionnaire closed with open-ended items focusing on VIPER®'s usability. These data were used to fine-tune maneuver accuracy and prepare the system for the evaluation.

Introduction Session and Scenarios

IPs were asked to develop a single-session introduction course to inform participants of the study and familiarize them with the ITDs prior to use. IPs also created nine scenarios utilizing the maneuvers developed in VIPER® where seven scenarios focused on the Contact Phase and two on the Instruments Phase of training. These scenarios provided details on both the mission and training objectives, suggested study reference, starting state of the aircraft, and maneuvers to be practiced.

Participant Binders

The binders were distributed to SNAs from both experimental conditions participating in the study. The only differences between the materials provided to each group was that the VIPER® participants received a VIPER®-specific user guide and the T-6B VIPER® questionnaire. The binder materials included the following:

- IRB participation documentation (the Privacy Act for review and the Informed Consent Document (ICD) to read and sign)
- Reference sheet for on-site device support
- Evaluation syllabus containing flight scenarios to practice. The scenarios employed the following VIPER® maneuvers[5]:
    - Takeoff
    - Power on Stalls
    - Approach Turn Stall
    - Landing Attitude Stall
    - GX
    - Steep Turns
    - Level Speed Changes
    - Landing Pattern
    - ILS Approach
    - Localizer
    - Unusual Attitude Recovery (VMC)
    - Unusual Attitude Recovery (IMC)
    - Slow Flight
    - Radial Intercepts
    - Arcing
    - Arc and Radial Intercepts
    - Constant Airspeed Climbs
    - Constant Airspeed Descents
    - Waveoff
    - Precautionary Emergency Landing (PEL)
    - Precautionary Emergency Landing in Pattern (PELP)

[5] Note: Only 26 of 33 developed maneuvers within VIPER® were used in order to have events most representative of actual syllabus events as well as leveraging the most developed and accurate maneuvers within the system.

  - o Turn Pattern
  - o Power Off Stall
  - o Aborted Takeoff
  - o Intentional Spin
  - o VFR Straight-In Approach

- Start-up guide for both the ITDs and VIPER® systems that included basic operating procedures (start-up, login/logout, set scenario parameters, navigate the system, and care for the system), troubleshooting instructions, and sanitizing procedures based on COVID-19 Command policy

- Logbook to track practice session start and end times, scenarios completed, repetitions of maneuvers, and issues encountered (see Appendix 7)

- T-6B VIPER® SNA Questionnaire for feedback on usability and utility (see Appendix 6)

T-6B VIPER® SNA Questionnaire

SNAs in the VIPER® condition, were given a 37-item questionnaire to provide feedback on their experience with the system (see Appendix 6). SNAs were asked to provide brief demographic information. SNAs provided feedback on the quality of the instructor-led overview on a 4-point Likert scale from “not helpful at all” to “extremely helpful.” Other items focused on effectiveness of VIPER® for their current Primary curriculum, 4-point Likert scale ranging from “not effective at all” to “extremely effective. There were also items regarding the effectiveness of each maneuver practiced using the same 4-point effectiveness scale. SNAs were also asked to provide their input on whether or not VIPER®’s feedback was timely, accurate, and informative. The survey concludes with items addressing VIPER®’s reliability, functionality, and ease of use on a 4-point scale ranging from “strongly disagree” to “strongly agree.” These data were collected to provide qualitative feedback about the system’s attributes.

T-6B VIPER® Wrap-Up Questionnaire

At the conclusion of the study, a 12-item questionnaire was emailed to IPs, stakeholders, and leadership to obtain their feedback on VIPER®’s overall potential and capability (see Appendix 8). The questionnaire consisted of free-response items divided into three sections: overall usability, coaching and feedback, and implementation. Items covered, but were not limited to the following topics: the benefits and limitations of the system, their experience with major components of the system, improvement in instruction, and improvement in SNA’s flight skills. These responses were used to identify trends about VIPER® as well as recommendations for improvements.

4.2.2 Apparatus

During the evaluation, four T-6B ITDs housed in a separate room from other virtual trainers were utilized for uploading the VIPER® program and data collection. The ITDs consist of desktop computers configured with head mounted displays, flight controls (control stick, throttle, and rudder pedals), flight simulator software, and a flight model of the T-6B Texan II aircraft. The same four devices were used for both the control and experimental conditions to practice the prescribed scenarios, see Image 1.

Image 1: T-6B ITD at NAS Corpus Christi, TX

4.3 Assumptions

It is assumed this study had no impact on the training schedule or the syllabus for the T-6B community. Performance data collected from aircraft training sessions were a part of

CNATRA's traditional grading and training feedback process. All data were delivered electronically via a secure mechanism from CNATRA to the NAWCTSD research team for data analysis purposes. Study participants practiced all of the scenarios developed for the evaluation and used the resources that were provided appropriately. Participants from the Assigned VR condition were not exposed to any VIPER®-related features and SNAs in the Archival condition had little to no VR experience during their Primary Training.

4.4 Procedures

4.4.1 Preparation for Data Collection

In preparation for data collection, DMI provided an introductory overview of VIPER® to IPs and stakeholders. From there the IPs provided a list of maneuvers that would be appropriate for practice within the system. DMI and CNATRA engaged in an iterative process for development, testing, and feedback. Once a validated list of maneuvers was delivered, the research team distributed the Curriculum Breakdown Survey along with the VIPER® Maneuver Feedback Questionnaire to be completed by IPs. Based on IPs' responses, additional development was required to fine-tune targeted maneuvers and system abnormalities experienced.

During this time, IPs created the introduction session as well as nine scenarios for practice on the ITDs with or without VIPER®. In parallel, the research team finalized measures and created participant binders to be distributed to each SNA at the start of the study. Scenarios mainly focused on maneuvers in the Contact Phase and the first few events of the Instrument Phase.

4.4.2 Study Design for Data Collection

A two-tailed G-Power Analysis was conducted with an effect size of 0.12, significance level of 0.05, and a power of 0.90. The power analysis revealed a recommended minimum number of 64 SNAs per condition (i.e., Assigned VR and VIPER®) executing 18 or more training hours in the ITDs to be able to detect medium-sized (approximately half a standard deviation) significant effect between groups. Assuming a class of 8-15 SNAs would enter into the study weekly, the research team planned to collect data from multiple classes.

To start the experimental portion of the study, CNATRA personnel provided introduction sessions of the ITDs for the VR condition. Sessions were conducted weekly, timed with the start of each new class of students entering Primary Training. At this time, SNAs were provided the Privacy Act statement and Informed Consent Document to read and sign. The SNAs were reminded that participation was voluntary and given contact information for any questions they may have about the study. CNATRA personnel also provided each participant with the data collection binders.

Next, SNAs were asked to complete the 18 practice hours in the ITDs over a 9-week period, working around their normal training schedule. It was estimated each prescribed scenario would take an hour to adequately complete, therefore, SNAs were encouraged to complete each of the nine practice scenarios twice to reach their 18 hours. Every week, SNAs were required to complete logbooks which were verified by CNATRA personnel and electronically delivered to the NAWCTSD research team.

Once all the Assigned VR SNAs were underway and VIPER® development was completed, CNATRA personnel provided introduction sessions for SNAs in the VIPER® condition. Similar to the Assigned VR condition, data collection binders were distributed, with the addition of a VIPER® startup guide to help SNAs access pre-developed scenarios and the questionnaire to allow SNAs to provide feedback on VIPER®'s utility and usability.

At the completion of data collection for each group, performance data from Contact and Instruments events were collected for comparison. Performance data were also obtained for the Archival and Free VR groups. Finally, IPs, leadership, and stakeholders who interacted with VIPER were invited to provide feedback via a wrap up questionnaire.

Analyses

For performance data, all analyses were conducted in IBM SPSS Statistics 26 for Windows (IBM, Armonk, NY) with default settings. A two-tailed alpha level of .05 was used for significance in all analyses. Due to violations of normality, violations of homogeneity of variance, and unequal sample sizes, Mann-Whitney U tests were used to conduct pairwise comparisons of the Archival, Free VR, Assigned VR, and VIPER® groups. Two effect sizes are also reported. As a nonparametric effect size related to Mann-Whitney U, the research team calculated η^2 on ranks, that is, the proportion of variability in ranks associated with group membership. In addition, to provide a clearer picture of the size of VR and VIPER®'s effect on performance, the research team calculated Hedges' *g*, that is, the difference between groups measured in standard deviations. However, Hedges' *g* in this report should be interpreted with caution, due to the violations of the normality and equal variance assumptions.

The primary performance comparisons between groups were the comparisons of grades (referred to as Event Raw Score, ERS). In each event, a number of maneuvers are completed, and each maneuver has a minimum required grade, known as the Maneuver Item File (MIF). ERS is calculated as:

$$\frac{Sum\ of\ Maneuver\ Grades}{Sum\ of\ MIF}$$

Thus, an ERS of 1 indicates adequate performance, less than 1 indicates poor performance, and greater than 1 indicates better-than-adequate performance. If VIPER® SNAs had better performance than the other groups, then their ERS should be higher. ERS was compared between the four groups for flights (4000-level events) in the Contact and Instruments Phases only. This is because IPs deemed VIPER® to be best suited for aiding SNAs at these stages and therefore designed scenarios to prepare SNAs for Contact and early Instruments events. Contact and Instruments are the first two phases of the syllabus that include live flights. Live flights were the focus of this evaluation because they represent the most critical measures of pilot performance. Events included in the comparisons are listed in Table 3 of (Appendix 1).

Beyond event grades, 4000-level events marked with various modifiers (adaptation sorties, practice sorties, warmup sorties, extra training, progress checkrides, repeats, and unsatisfactory events) were also compared between groups for Contact and Instruments events, as higher numbers of these events can serve as an indicator of worse performance or reduced training efficiency. However, counts of modified events were unavailable, as the data received only contained the final instance of each event and did not include multiple iterations. For example, if an SNA completed event C4101 three times, then the event was repeated twice, but only one repeat (the third/last attempt) would be recorded in the data file. Therefore, in order to approximate the relative frequency and evaluate potential group effects, the percentage of events marked with each modifier code was calculated for each SNA. Additionally, by employing percentage rather than raw counts, it accounted for variation in the number of events completed and recorded for each SNA.

Finally, the research team counted the number of participants who received at least one "Unsatisfactory" rating compared to those who received none. These counts were compared between groups using Fisher's Exact Tests. As with ERS, comparisons were limited to the Contact and Instruments phases.

For feedback data, the research team summarized responses to both the SNA and wrap-up questionnaires. Due to low response rates and the fact the SNA questionnaire was administered to SNAs who used different versions of VIPER®, statistical analyses of ratings were not conducted, other than median and interquartile range for some specific responses of interest. As a result, the team focused largely on identifying feedback trends and highlighting recommendations from the qualitative data.

It is important to note these analyses included participants who used VIPER® in its initial operational state. DMI further developed VIPER® based on feedback from the SNAs, including updates to increase reliability and address frequent system crashes resulting from interactions with other ITD software programs and updates. Thus, performance and feedback results may be less strongly positive than they would be in a future analysis in which the more reliable version of VIPER® was the only version used.

5. Results

The research team felt it important to explain a few limitations in interpreting the data prior to the discussion of the results to allow for better understanding of the findings below.

- According the IP focus group, students in the VIPER® condition may have completed some of their practice hours without VIPER® enabled, making practice similar to those in the Assigned VR group. This would introduce an unanticipated confound in this evaluation by reducing differences between the VIPER® and Assigned VR conditions.

- High study attrition rates in the Assigned VR and VIPER® conditions may indicate that only highly motivated and high performing students completed the study in these groups. This may have affected results, such that Assigned VR and VIPER® performance appear higher than it would be with a more representative sample of students.

- Frequencies of event modifier codes could be attributed to situational factors like scheduling and weather, making it difficult to identify the true cause for the additional flights. Unsatisfactory events (unsats) are the most directly tied to student performance, with fewer unsats indicating better performance.

- Statistical significance is often based on the traditional *p*-value of < .05. However, it should be noted that due to the number of comparisons examined (over 200), it can be expected a small portion of significant results (5% or approximately 10 comparisons) are Type I errors (i.e., false positives). For more robust conclusions, more stringent *p*-values were applied to individual event comparisons and differing *p*-values are denoted in the tables (* = $p < .05$, ** = $p < .01$, + = $p < .001$).

5.1 Performance

Six Mann-Whitney *U* tests were conducted for each variable in the Performance section, one to compare each of the four groups to each of the other groups. The only exceptions were some event modifiers in which one or more groups did not have any of the event modifiers being compared. These exceptions included unsats, the modifier code most closely associated with performance; VIPER® SNAs did not have any unsats during Contact or Instrument flights and Assigned VR SNAs did not have any unsats in Instruments flights. Therefore, unsats were compared between groups using Fisher's Exact Tests. SNAs who completed at least one flight in the relevant phase(s) were included in analysis.

As mentioned previously, VIPER® scenarios focused heavily on skills related to Contact events, with a few early Instruments skills included as well. As a result, the research team focused performance analyses on Contact and Instruments events and expected to find significant differences between the VIPER® and VR groups in Contact events. Significant differences in Instruments events were not expected to be as prevalent due to the relatively small emphasis on Instrument-specific skills.

5.1.1 Event Raw Score

For each SNA, ERS was compared separately for each live flight in the Contact and Instruments phases. Two flights, C4501 (the initial solo flight) and C4801, were not graded often enough to be compared between groups, so these two flights were excluded from flight-by-flight comparisons; all other Contact and Instruments flights were included. In addition, the research team calculated the mean ERS across all Contact and Instruments flights, across Contact flights only, and across Instruments flights only. Because the research team did not have detailed maneuver-level data for each event, but only ERS, mean ERS was calculated as:

$$\frac{Sum\ of\ flight\ ERSs}{Number\ of\ flights}$$

This differs from CNATRA's method of calculating overall grade, which uses the sum of grades divided by the sum of MIFs across all graded events. See Table 1 for overall Contact and Instruments results, and Table 3 in Appendix 1 for results of individual event comparisons. Significant ($p < .05$) Mann-Whitney *U* test results are marked with superscripts. In addition, rows with significant results are marked with bold text.

Table 1. Mann-Whitney *U* tests on overall Contact and Instruments phase ERS

Event	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
All Contact + Instruments Flights	Archival	1.13 (0.05)	836	-	-	-
	Free VR	**1.16 (0.07)**	**3014**	**917,589.5^{+}**	**.038**	**0.42**
	Assigned VR	**1.18 (0.08)**	**64**	**60,082.5^{+}**	**.009**	**0.33**
	VIPER®	**1.23 (0.08)**	**52**	**1099****	**.086**	**0.60**
All Contact Flights	Archival	1.15 (0.06)	836	-	-	-
	Free VR vs	**1.18 (0.07)**	**3014**	**933,542.5^{+}**	**.034**	**0.41**
	Assigned VR	**1.20 (0.09)**	**64**	**65,487.5^{+}**	**.006**	**0.34**
	VIPER®	**1.24 (0.08)**	**52**	**1264***	**.043**	**0.49**
All Instruments Flights	Archival	1.10 (0.03)	836	-	-	-
	Free VR	**1.11 (0.04)**	**2506**	**817,064^{+}**	**.015**	**0.29**
	Assigned VR	**1.16 (0.05)**	**64**	**38,414^{+}**	**.018**	**0.99**
	VIPER®	1.18 (0.08)	36	919	.022	0.34

Note. *M* and *SD* = mean and standard deviation, *n* = number of participants included in the Mann-Whitney *U* test, *U* = Mann-Whitney *U* statistic, η^2 = effect size for Mann-Whitney *U* test (proportion of variation attributable to difference in ranks), g = Hedges' g (difference between groups in standard deviation units). Mann-Whitney U tests, η^2, and Hedges' g are included for the comparison to the previous condition(i.e., the Free VR row shows the change from Archival to Free VR); the Assigned VR row shows the change from Free VR to Assigned VR, and the VIPER® row shows the change from Assigned VR to VIPER®. Detailed comparisons between all groups, broken down by event, are presented in Table 3, Appendix 1. Significant Mann-Whitney U tests are indicated with bold text. * = $p < .05$, ** = $p < .01$, $^{+}$ = $p < .001$.

Differences in Grades

Across event comparisons, the general pattern of performance was that VIPER® SNAs had the highest ERS, followed by Assigned VR, Free VR, and finally the Archival group received the lowest grades. The complete pattern occurred in 30% of events (42% of Contact events and 14% of Instruments events), although not all differences were significant. An additional 33% of events (42% of Contact events and 21% of Instruments events) mostly followed the same pattern, but with one comparison in which the groups had equal performance (Archival = Free VR, Free VR = Assigned VR, or Assigned VR = VIPER®). Six events (C4602, C4790, I4102, I4103, I4104, and I4203) showed slight decreases in ERS for VIPER® compared to Assigned VR, but none of the differences were significant. The overall pattern appeared to mostly support the benefits of assigned VR training with some additional benefit from demonstration and deliberate practice with feedback provided by VIPER. See Figure 1 for an illustration of ERS differences across individual flight events.

Statistical comparisons of individual events support the pattern of improved performance across the four groups where significant differences were found between each consecutive level. However, since a substantial number of comparisons were completed (i.e., 3 group comparisons x 33 events = 99 tests), the results in the section should be interpreted cautiously due to an increased risk of false positives; the false positive criterion rate (i.e., p-value or α) was adjusted to .01 to mitigate this risk but is not a complete solution. Bearing that caveat in mind, the Free VR condition had a significant overall increase in ERS compared to the Archival group in 76% of events (84% of Contact events and 64% of Instruments events). The Assigned VR condition had a significant overall increase in ERS compared to the Free VR group in 42% of events (47% of Contact and 36% of Instruments). Finally, the VIPER® condition had a significant increase in ERS compared to the Assigned VR group in only one Contact event, C4304 (the last Day Contact event before the

Midphase Contact Checkride). Although differences between VIPER® and Assigned VR were not as broadly prevalent as differences between Assigned VR and earlier conditions, VIPER® still provided benefit beyond the Assigned VR condition. Overall, these comparisons indicate the strong value of making VR trainers freely available to SNAs, allowing demonstration of maneuvers, providing structure to the practice in VR trainers, and ensuring feedback is provided to guide SNA practice.

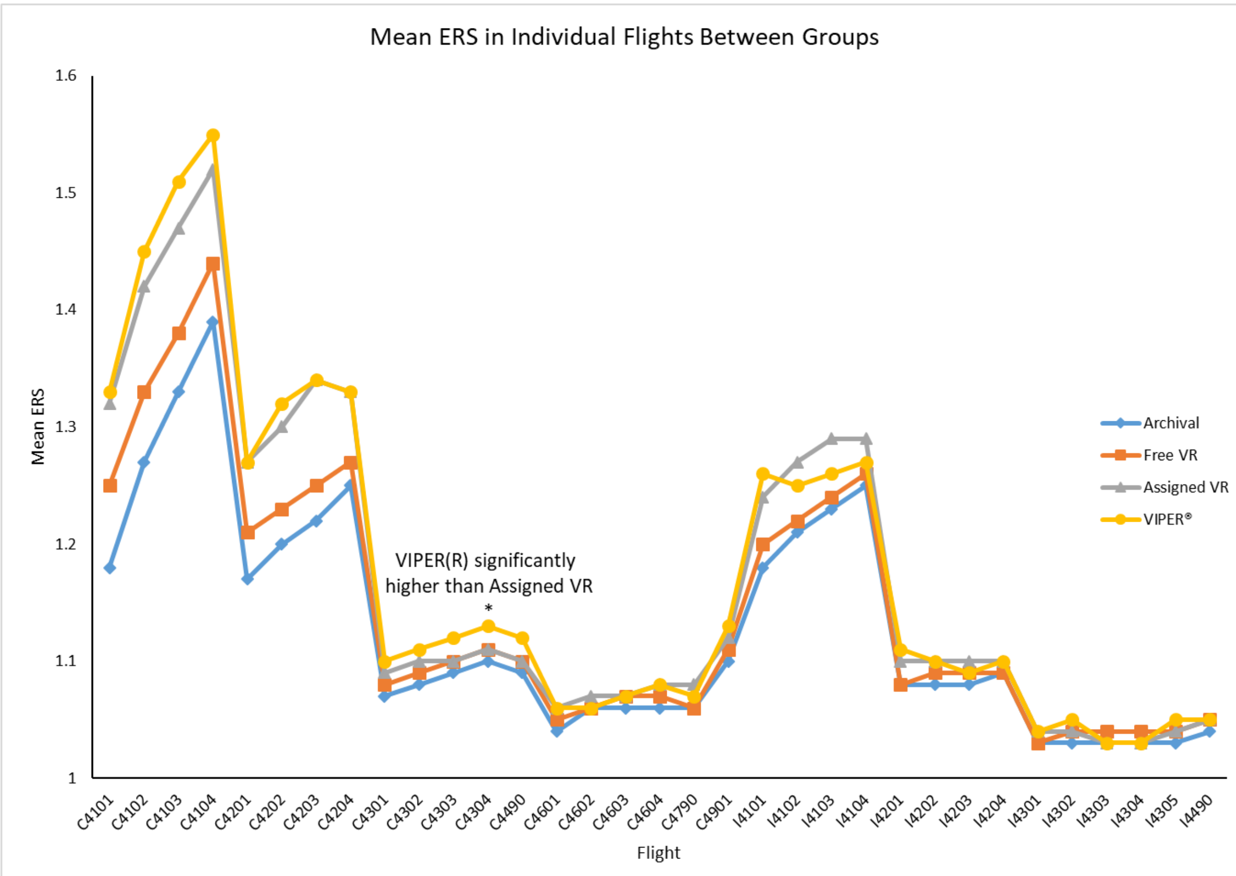

Figure 1. ERS in individual flights.

Note. The significant difference ($p < .01$) is indicated in Figure 1 for the VIPER® vs. Assigned VR conditions. For all significant differences, refer to Table 3 of Appendix 1. Error bars are excluded from this graph for visual clarity.

Results are arguably even stronger when examining at the aggregate levels combining events into three categories: 1) an overall of all flights examined, 2) only contact flights, and 3) only instrument flights. This also shows each group with consecutively higher ERS: Archival remains the lowest grades and

VIPER® the highest. All but one comparison show statistically significant differences ($p < .05$), instrument flights between Assigned VR and VIPER®, likely due to the greater emphasis on the Contact Phase for scenarios and VIPER® maneuver development as well as small sample sizes for Instrument flights among the VIPER® condition. See Figure 2 for an illustration of aggregate results.

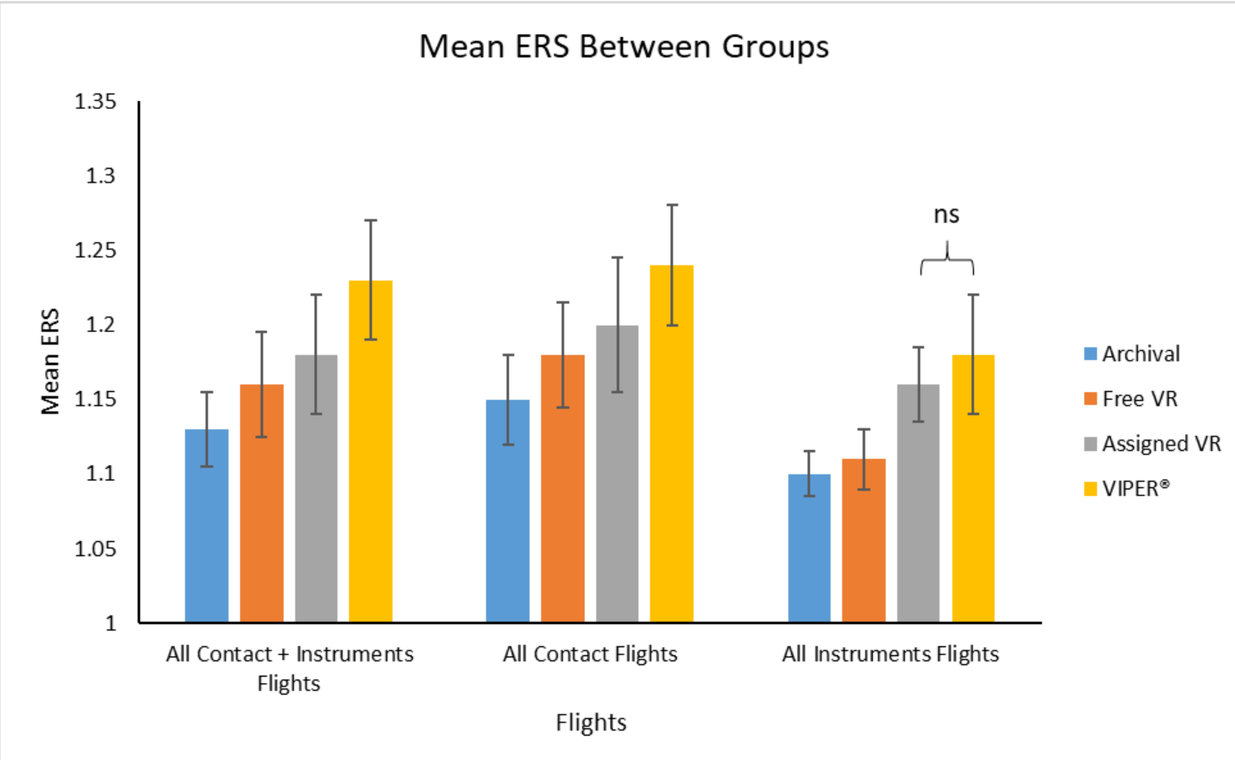

Figure 2. Aggregate Mean ERS Between Groups.

Note. Error bars indicate ½ standard deviation above and below the mean. All differences are significant ($p < .05$) except for the difference between Assigned VR and VIPER® ERS in Instruments flights.

Magnitude of Effect

Examining the effect sizes (Hedge's *g*) can show the impact demonstration, deliberate practice, and feedback provides. Though it should be interpreted with caution due to the non-normal nature of the data, and why it is only discussed here for the Overall ERS, it can provide an approximation of the training effect. Between Archival and Free VR, the effect size is 0.42, indicating nearly half a standard deviation improvement in ERS. The difference between Free VR and Assigned VR finds a 0.33 effect size increase in grades and then from Assigned VR to

VIPER® shows a 0.60 effect size. Translating what that means into Naval aviation training grades where a standardized T-score (Navy Standardized Score: NSS) is used with a mean of 50, standard deviation of 10, with lower-bound of 20 and higher-bound of 80, an SNA could see an increase of approximately 5 (ranging 3.3 to 6) NSS points by utilizing VIPER® beyond just completing assigned VR practice. Larger still would be the increase if compared to no VR (archival) or Free VR (no assigned practice/scenarios)[6].

Variability in Rank Orders

Further analysis of significant event comparisons examining the proportion of differences in ranks that can be attributed to group membership found η^2 ranged from .002 to .105, indicating that between 0.2% and 10.5% of the variability in ranks is associated with the level of demonstration, deliberate practice, and feedback (i.e., study condition). Interestingly, the 10.5% difference occurs between the Assigned VR and VIPER® groups for the comparison of C4304 and a 5.6% difference occurs for C4303, the two flights before the Contact checkride and solo. These relatively large differences combined with considerable Hedges' *g* values indicate a strong advantage of having used VIPER® leading up to the solo flight and provide further evidence for the importance of demonstration, deliberate practice, and feedback.

5.1.2 Event Modifier Codes

For each event modifier code (adaptation sorties, practice sorties, warmup sorties, extra training, progress checkrides, elimination checkrides, repeats, and unsatisfactory events), the percentage of Contact and Instruments flights that included a modifier was calculated for archival, VR, and VIPER® SNAs. In some cases, no SNA in a group had any event with a given modifier; these all-zero instances are indicated in Table 4 of Appendix 2. Mann-Whitney *U* tests could not be conducted on groups with all-zero counts. Importantly, VIPER® SNAs did not have any unsats, and Assigned VR SNAs did not have any unsats in

[6] Reported in Table 3 in Appendix 1, the Hedges' *g* effect sizes for these comparisons are not discussed here as it was felt the single group difference estimates were more accurate due to being closer in time of assessment. For example, the Archival group went through training prior to October 2018 and the VIPER® condition occurred in 2021. Though training should remain relatively stable, there can be fluctuations, so only subsequent pairs are discussed.

Instruments flights, possibly indicating that their ITD practice reduced the chance of unsatisfactory performance. All other percentages were compared using Mann-Whitney *U* tests; see Table 2 and Figure 3 for significant results, and Table 4 in Appendix 2 for all results.

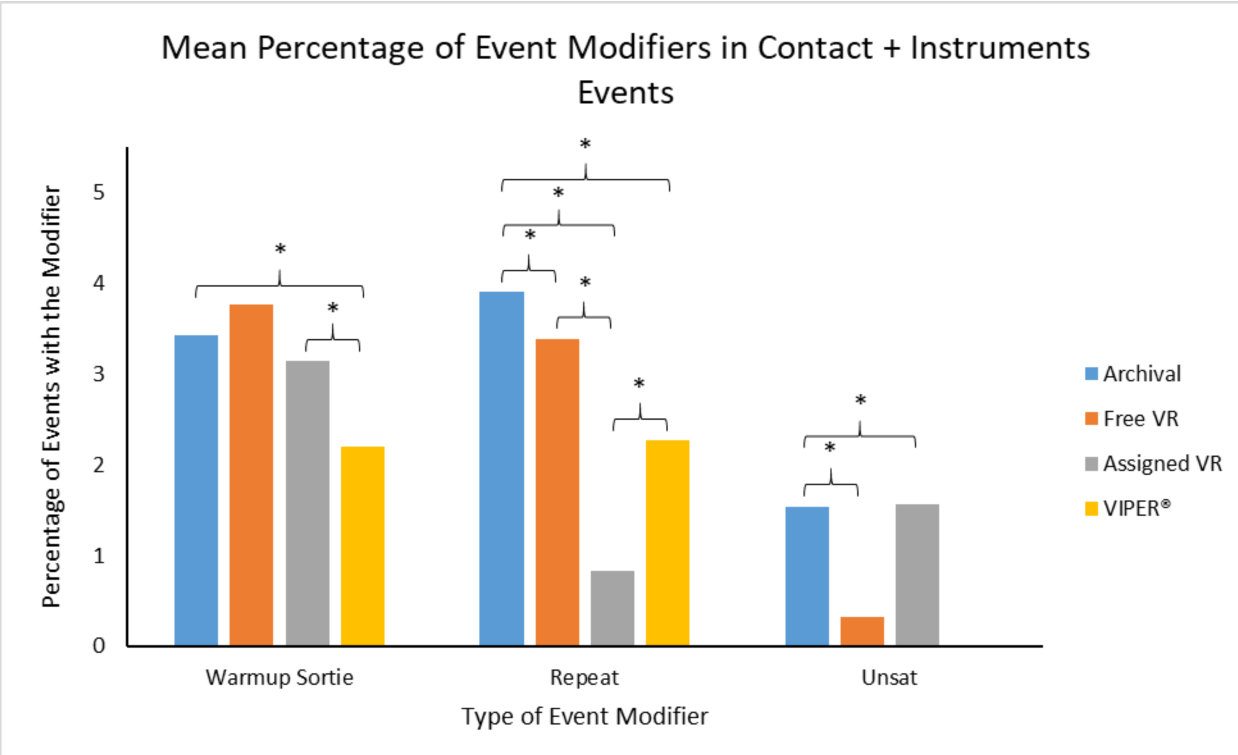

Figure 3. Percentage of events with warmup sorties, repeats, and unsats.

Note. The VIPER® unsat bar is not visible because VIPER® SNAs had no unsats. Significant differences ($p < .05$) are indicated with asterisks. Error bars are excluded from this graph for visual clarity.

Results were not as strong as ERS comparisons, but in general followed a similar pattern: the VIPER® and Assigned VR conditions had the lowest percentage of events with modifiers, followed by Free VR, and finally the Archival group, on average, had the highest percentage of events with modifiers. The differences between Assigned VR and VIPER® did not show a clear advantage of one over the other; the Assigned VR group had more warmup sorties, but the VIPER® group had more repeats. Thus, being assigned to practice in a VR device, with or without a virtual instructor, was associated with a reduced percentage of event modifiers. These results should be interpreted with some caution since reasons for event modifiers can vary from weather cancelations to resource limitations to inadequate performance and, without data on reasoning, it is difficult to draw robust conclusions (with the exception of unsats).

Table 2. Significant results from Mann-Whitney *U* tests on percentage of event modifiers

Event Modifier	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
Warmup Sortie (Contact + Instruments Flights)	Archival	3.43% (4.39%)	836	-	-	-
	Free VR	3.77% (6.76%)	3014	1,227,788.5	<.001	0.05
	Assigned VR	3.14% (3.55%)	64	92,400	<.001	-0.09
	VIPER®	**2.20% (3.81%)**	**52**	**1317.5***	**.037**	**-0.26**
Repeat (Contact + Instruments Flights)	Archival	3.91% (4.60%)	836	-	-	-
	Free VR	**3.39% (5.42%)**	**3014**	**1,117,509$^+$**	**.007**	**-0.10**
	Assigned VR	**0.83% (1.81%)**	**64**	**66,761.5$^+$**	**.007**	**-0.48**
	VIPER®	**2.27% (4.03%)**	**52**	**1375***	**.037**	**0.48**
Unsat (Contact + Instruments Flights)	Archival	1.53% (5.13%)	836	-	-	-
	Free VR	**0.32% (3.51%)**	**3014**	**964,725.5$^+$**	**.125**	**-0.31**
	Assigned VR	1.56% (12.50%)	64	95,181.5	<.001	0.32
	VIPER®	N/A: VIPER® condition did not have any unsats				

Note. This table includes only Warmup, Repeat, and Unsat results for Contact + Instruments Flights; detailed results broken down by phase and additional event modifier codes are included in Table 4 (Appendix 2). *M* and *SD* = mean and standard deviation, *n* = number of participants included in the Mann-Whitney *U* test, *U* = Mann-Whitney *U* statistic, η^2 = effect size for Mann-Whitney *U* test (proportion of variation attributable to difference in ranks), *g* = Hedges' *g* (difference between groups in standard deviation units). Significant Mann-Whitney *U* tests are indicated with bold text. * = $p < .05$, ** = $p < .01$, $^+$ = $p < .001$.

With unsats as a measure of a SNA's performance but unable to conduct Mann-Whitney *U* tests for significance, the research team compared the groups' odds of receiving at least one unsat in the Contact and Instruments Phase flights. Separate Fisher's Exact Tests were conducted for the Contact Phase, Instruments Phase, and Contact + Instruments combined. All three omnibus Fisher's Exact Tests (testing on all four groups at once) were significant (p < .001 for all), indicating significant group differences in the odds of having at least one unsat. Post-hoc

Fisher's Exact Tests (comparing two groups at a time) indicated that the differences were driven by the Archival group. In the overall (Contact + Instruments) and Contact Phase comparisons, the Archival group had significantly higher odds of receiving an unsat than the other three groups; in the Instruments Phase, the Archival group had significantly higher odds than the Free VR group ($p < .001$ for all). All other comparisons were non-significant. SNAs in the Archival group were 12-38 times more likely than other groups to receive at least one unsat. These results are presented in Table 5 (Appendix 3).

5.2 SNA Feedback

The T-6B VIPER® Questionnaire was administered to SNAs who completed their allotted 18 hours or more in VIPER®. Responses were received from 15 out of 52 SNAs. The results of the questionnaire are not reported in full because they included feedback from SNAs who used VIPER® at different operational states. Only some notable responses are provided in this report.

5.2.1. Utility

Four general results on utility are of note.

1. On the question about what VIPER® could be used to accomplish, SNAs most frequently responded that it was useful for preparing for an upcoming event (80%), followed by building a sight picture (53%); (20%) SNAs also wrote in the free-response option that it could be used to learn course rules.

2. Almost all SNAs (93% for each possible response) believed skills from every chapter of the syllabus (Contact, Instruments, Navigation, and Formation) could be practiced on VIPER®.

3. Some maneuvers were not practiced by any SNA, those maneuvers that were practiced hovered around the middle of the scale when SNAs were asked to rate their effectiveness in VIPER®, with all median rankings falling between 2 ("slightly effective") and 3 ("very effective"). Thus, SNAs tended to agree that VIPER® could provide some effective practice on the maneuvers they experienced.

4. The majority of SNAs (64% to 92%, depending on the question) responded that VIPER® provided timely, accurate, and informative instructions and feedback.

These results indicate a generally positive response to VIPER®'s usefulness among SNAs, with one SNA stating that VIPER® made them a better student. One exception, indicated by three comments, was that VIPER® had trouble monitoring the aircraft's position in the landing pattern. SNAs were evenly split (7 "yes" and 7 "no" responses with 1 non-response) on whether or not they would recommend VIPER® to future SNAs. However, this must be interpreted in light of the fact that some responses came from SNAs who used VIPER® in its original state; the proportion of "yes" responses may have been higher if all SNAs had used the updated version of VIPER®. Among those who would recommend it, three commented that future SNAs would benefit from using VIPER® to build a sight picture of the maneuvers.

5.2.2. Usability

Overall, SNAs considered VIPER® easy to use. They found the text and auditory stimuli to be clear and understandable (Median = 3 out of 4, "Agree"); and most did not find VIPER® distracting (Median = 2 out of 4, "Disagree"). Software crashing, as expected, was an issue (Median = 2 out of 4, "Disagree," when asked if VIPER® could be relied on not to crash), but was the only usability issue consistently noted in the questionnaires. Crashing was initially very frequent due to an unforeseen change in software on the VR devices, but was addressed by DMI after initial feedback. Three SNAs also commented that VIPER® did not always register or understand when they asked questions, a result that may have been affected by ambient noise within the building housing the VR devices, or by limitations in voice recognition software.

5.3 IP Feedback

At the conclusion of the study, a 12-item, questionnaire was emailed to IP, stakeholders, and leadership to obtain their feedback on VIPER's overall potential and capability. However, due to their limited involvement during the study, we were not expecting high response rates. As expected, survey responses were received from the two IP most involved and familiar with the VIPER effort at NAS Corpus Christi. Their responses are summarized below.

5.3.1 Utility

Overall, IPs reported some value in VIPER® as an early training tool. However, they indicated some limits to VIPER®'s utility, especially for maneuvers that require flexible grading like the landing pattern, which can be completed successfully in multiple ways. In addition, they made suggestions for improving VIPER® in the future, including additional maneuver development, a different grading scale, and grading of partially completed maneuvers.

Generally, IPs considered practice with VIPER®'s maneuvers to be useful for self-study prior to flying and capable of honing skills for some maneuvers. It was stated that SNAs who saw VIPER® as beneficial and made an effort to learn how to use it improved their performance, although it was not specified how performance improved.

The verbal feedback provided by VIPER® was described as "focused and timely" and therefore useful when it was accurate. It was specifically called out as a potential help for SNAs to build scan patterns. However, the IPs mentioned two limitations to the verbal feedback:

1. It did not provide theoretical information: it explained how to complete a maneuver, but not why it should be completed that way.

2. It was not accurate for all maneuvers, especially for maneuvers that require more flexibility in grading (i.e., those maneuvers that are often completed in a non-standard but acceptable manner).

Another comment also stated that inaccurate grading for some maneuvers could likely be corrected by conducting a more extensive review with a larger number of IPs, but expected some maneuvers to require too much subjective judgment in grading to be handled well by VIPER®'s current grading structure.

Regarding the instructor dashboard, it was not known if anyone had actively used the dashboard, but from their review, the IPs offered a few suggestions. It was indicated the type of information it provided was useful, but would be more useful with two changes: 1) present results in terms of percent of maneuver correctly completed rather than in an arbitrary points-based scale; and 2) record and present information for maneuvers partially completed either through intentionally ending an event early or through a system crash. IPs suggested that recording and presenting partially completed maneuvers may also decrease any frustration associated with system crashes.

Instructors also expressed concerns about the A2A flight model, which is a part of the T-6B ITD itself, and not a part of VIPER®. However, it was noted that VIPER® would not likely be useful for training aerobatic maneuvers until the ITD's flight model becomes more accurate. This suggests potential greater utility for VIPER® or similar programs when it is used in conjunction with more realistic subsystems.

The IPs agreed VIPER® is best early in training, especially prior to entering Primary flight training. VIPER® was deemed to be much more valuable when used after the Naval Introductory Flight Evaluation (NIFE) program, when it could serve as an introductory tool prior to entering Primary training in which SNAs have access to the ITDs for pre-flight practice. Additionally, IPs deemed it better logistically to use VIPER® as an early training tool instead of during Primary training because:

1. SNAs have less to no need of VIPER®'s instruction once they have started flying the actual aircraft;

2. SNAs already have a full schedule during Primary training; therefore, VIPER® could potentially interfere with training by drawing the SNA's resources away from items in the existing Primary training syllabus.

5.3.2 Usability and Operability

With regard to operability, IPs reported that software crashing was a primary concern, and though updates to improve reliability sufficiently addressed the issue, lingering mistrust of the system still remained. In addition, although IPs did not use this function, they stated VIPER®'s interface for creating and editing event profiles appeared easy to use.

Usability issues associated with voice recognition were noted. IPs reported that it was often necessary to repeat a command multiple times to be understood by the system. More critically, IPs estimated that 20-30% of the time there was a disconnect between command and action where VIPER® started an unrelated maneuver to the one requested, and not necessarily with a similar-sounding name. IPs reported this was so frustrating that students would often give up on practicing the originally-intended maneuver; IPs also stated that the disconnect between commands and resulting actions damaged SNA users' perception of VIPER®'s utility, which harmed its reputation among all SNAs and decreases willingness to use VIPER®.

5.3.3 Reception among IPs and SNAs

Though several positive reports from a few IPs and some SNAs were received, overall reception to the use of VIPER® as a training tool was slightly negative as reported in the questionnaires. On the positive side, the IPs reported that some SNAs expressed value in using VIPER® and that they would recommend it to other SNAs. However, general reception among SNAs was primarily negative as communicated to IPs from students in the VIPER® condition and documented in student responses to the questionnaire. Early on, SNAs found the software crashes frustrating; and later, they continued to be frustrated by other program issues such as voice recognition failures and attributed any ITD system crash to the VIPER® program. The frustration led participants to speak disparagingly of VIPER® to classmates, creating a negative reputation even among SNAs who had never used it. Instructors reported that to some extent, SNA reception varied by class, with some classes disregarding requests to participate in the VIPER® evaluation and other classes considering it their duty to participate, but most SNAs did not expect VIPER® to be beneficial and therefore did not want to dedicate time to using it. IPs reported that they observed some students in the VIPER® condition not using the system during practice sessions, therefore, some of the practice hours SNAs logged for VIPER® may actually have been hours spent simply using the ITDs without VIPER® enabled in an attempt to avoid software crashes. Finally, the IPs stated that other instructors were not aware of VIPER®, were too busy to try VIPER®, or simply were not interested in learning about VIPER®. IP lack of interest and knowledge made it more difficult to curtail SNA frustrations and misunderstandings of the system, and further contributed to the less than positive reception.

6. Discussion

Overall, the research team expected the use of demonstration, deliberate practice, and feedback to increase SNA performance such that the Archival group would have the lowest performance, then Free VR, then Assigned VR, and the VIPER® group would have the highest performance. This is in line with theories surrounding expertise development and the learning sciences. The results largely supported these expectations, especially in the primary performance measure (grades). On the individual event level, the majority of Contact flights and several Instruments flights showed significant differences in the expected directions, although only one Contact flight showed significant differences for the virtual instructor added condition: Assigned VR vs. VIPER®. At an aggregate level, results were similarly strong where significant differences were seen between all conditions for the average ERS across all flights (Overall), only Contact flights (at $p < .05$), and nearly all Instrument flight comparisons with the only non-significant results being the Assigned VR vs. VIPER® conditions.

6.1 Grades and Event Modifiers

As expected, there were performance benefits across the four levels of VR usage; VIPER® slightly outperformed Assigned VR (no VIPER®), which outperformed Free VR (access to ITDs but no requirement to use them), which outperformed Archival (no access to ITDs). These differences occurred for both Contact and Instruments Phases combined as well as each phase analyzed separately (with the exception of the Instruments phase comparison between Assigned VR vs VIPER® which was not significant). When comparing individual flights, a large number of both Contact and Instruments events showed statistically significant differences between the Archival, Free VR, and Assigned VR grades; as well as between Archival, Free VR, and VIPER®, grades.

Though the results comparing VIPER® vs Assigned VR showed only minor improvement, the aggregate comparisons found significant increases in performance for the Overall average score and the Contact flights average, but not the Instruments average. However, only one individual events (i.e., C4304) showed significant gains for VIPER® above the Assigned VR condition. There are several likely contributing factors for this result:

1. The VIPER® maneuvers and prescribed VR scenarios were largely focused on the Contact-Phase, so any gains in performance would be expected to appear stronger during these events than Instruments.

2. Due to significant attrition, the VIPER® condition did not meet the sample size as determined by the power analysis, leading to somewhat lower statistical power than planned. Additionally, statistical power was reduced further in the Instruments phase than the Contact phase because most VIPER® SNAs did not completing the entire Instruments phase before data collection ended.

3. It is probable that statistically significant differences between VIPER® and other conditions could have been demonstrated across more events had a larger VIPER® group participated. For example, despite similar ERS between the groups, there were fewer significant results for the Archival-vs-VIPER® comparisons than Archival-vs-Assigned VR, indicating that the larger size of the Assigned VR group allowed for more robust detection of differences.

4. Due to the number of statistical comparisons conducted, familywise error rate increased and power was reduced potentially preventing identification of significant results. Though this was addressed by imposing a more conservative p-value criterion, it does require some caution in interpreting results.

5. The frequent crashing of the initial VIPER® system may have washed out benefits that would have appeared if all SNAs had been able to begin with VIPER® in its updated state.

6. With the accumulation of practice hours occurring as SNAs moved through the syllabus (i.e., little use of VIPER® early on in the syllabus), SNAs might not have received enough practice with VIPER® to show skill transfer for the early Contact events.

7. Though the prescribed scenarios were developed to be useful throughout the Contact phase and the beginning of Instruments, they may not have focused heavily enough on the later, more advanced Contact and Instruments events to lead to yield statistically significant differences in

performance. If so, this explanation would provide support for students needing to practice maneuvers numerous times prior to the flight events that are supported for best training outcomes. For example, injecting VIPER® in Ground School may be worth exploring for future integration options to increase the opportunities for practice before flying.

Of interest, C4304 was the only individual event to show strong benefits of VIPER® over Assigned VR at an alpha level of .01. This event is the last flight prior to the Contact Phase checkride and initial solo flight, which is a culmination of the stage. Unfortunately, the checkride did not show a statistically significant difference and the initial solo flight itself could not be tested since an IP is not part of the flight (it is the SNA's solo flight) and it is treated as a pass/fail event. This could be due to the cumulative effect of hours within VIPER® by the time SNAs reach the C43 events as well as VIPER® providing practice opportunities on 13 of the maneuvers graded in this block. Additional analyses with improved power would be required to determine if a difference truly exists.

Though it is possible the effects observed for the VIPER® group on C4304 (and others that were close to significance like C4303) were due to some factor outside the study, but when examined with the aggregate results, it appears highly unlikely. When ERS is averaged across all Contact events, the VIPER® condition had a statistically higher average grade ($p < .05$) than Assigned VR. Also, despite IP feedback stating some of the logged VIPER® practice times might have been ITD practice without VIPER® (same type of practice as the Assigned VR condition), not all VIPER® hours logged were actually VR-only hours. Therefore, if the VIPER® condition did contain what would actually be Assigned VR-type of hours, the difference between the two conditions may have been even greater since there was a statistically significant difference with potential condition cross contamination.

Examining the Instrument events, results were less strong but not entirely unexpected. Though VIPER® SNAs showed a trend toward slightly lower mean ERS than Assigned VR SNAs on several of the Instruments flights, none of these differences approached significance, indicating approximately equal performance for the two conditions. Additionally, when examining the mean across all

Instruments flights, VIPER® trended toward outperforming Assigned VR, although this result was not statistically significant. There are two likely contributing factors: one, VIPER® maneuvers and prescribed scenarios were largely Contact-focused so received a greater amount of demonstration, deliberate practice, and feedback than Instrument maneuvers or scenarios; and two, the small VIPER® sample sizes across the Instruments events did not provide enough power to determine any differences. Further data collection would be needed to determine if this pattern is real or spurious and additional maneuver and scenario development may be warranted to improve opportunities for demonstration, deliberate practice, and feedback on Instruments events.

Finally, the results from event modifier codes largely followed the same general pattern as grades. In the significant comparisons, VIPER® and Assigned VR SNAs had a smaller proportion of event modifiers than Free VR SNAs, who had a smaller proportion of event modifiers than Archival SNAs. The difference between VIPER® and Assigned VR was less clear: VIPER® SNAs had fewer warmup sorties, but Assigned VR SNAs had fewer repeats. In terms of the odds of receiving an unsat (a judgment of unsatisfactory performance in an event), SNAs in the Archival group were much more likely than the other groups to receive an unsat in at least one flight (12-38 times more likely, depending on the comparison). Among the event modifier codes, unsats are the most clearly related to performance, and therefore potentially the best modifier code indicator of an advantage for VR access, with or without a virtual instructor. For the most part, the results indicate a benefit towards using VR and ITSs not only for performance, but also potentially for training efficiency. However, the results from event modifier codes should be interpreted with caution. The differences in receiving an unsat between Archival and other groups are so extreme that they may indicate a factor other than the presence of VR is attributing to differences; for example, a cultural change over time whereby the tendency to judge an event as Unsatisfactory has decreased. In addition, the other modifier codes are not as closely linked to performance as unsats, and therefore may also be attributed to changes not associated with SNA performance.

Overall, the results provide strong support for the notion that demonstration, feedback, and deliberate practice in VR devices is worth the investment for the aviation training community.

Translating the effect sizes into NSS, the overall performance metric used in Naval aviation training, sees approximately 3-5 point increases for each subsequent level, a non-trivial amount especially when NSS is used to determine eligibility for certain training pipelines. Additionally, though not as strong, there was indication that VR and potentially ITSs reduce flights with event modifiers, improving training efficiency and reducing costs. When combined (i.e., grade and event modifiers), these results show the advantage in utilizing VR systems for practice can have on training outcomes and lend support for the use of VIPER® or other ITSs as a way to encourage deliberate practice and provide theory-based process and outcome feedback to further enhance performance. These results are especially important for the Military community, in which VR is being rapidly incorporated into training pipelines with limited objective data to guide acquisition decisions.

6.2 SNA Feedback

Although the team did not receive responses to utility and usability of the virtual instructor from the majority of the students who participated in the study, the feedback that was provided expressed that VIPER® was useful for building sight pictures and preparing for events. This comment has been consistently mentioned when ITDs are available for practice (McCoy-Fisher et al, 2019). These devices with the added benefit of performance feedback can be invaluable to learning by allowing students hands-on exposure to the aircraft. It is also worth noting that even though VIPER® only included maneuvers from the Contact and Instruments phases, SNAs did report the system could support all phases of Primary Training if fully developed to do so.

In reference to usability, students reported VIPER® was satisfactory for its current state, but could use further development. They stated it was not distracting and easy to use once beyond the crashing issues present with the initial version. Students found written and auditory feedback to be clear, but a few had issues with the system understanding commands and questions. These comments could be related to noise in the environment, limitations of speech recognition software, or the initial VIPER® crashing issue. Refinements in development could alleviate or improve on these areas.

6.3 IP Feedback

The two IPs who responded to the Wrap-Up Questionnaire reported some value in VIPER®, particularly as an early trainer to be used for self-study, indicating post-NIFE but pre-Primary usage as its most likely advantageous application. They thought it helped build scan patterns and hone flight skills for some maneuvers, especially maneuvers in which grading does not need to be highly flexible. On the more flexible maneuvers, such as landing pattern, it was not accurate enough to be considered beneficial. However, they felt that utility could be increased with changes to performance recording and feedback and by pairing VIPER® with a simulator that contains a more accurate flight model than the current T-6B ITDs. Recommendations based on their comments are included in section 6.4, “Recommendations.” According to IPs, after software crashing was addressed, the main remaining usability issue was trouble with voice commands: commands sometimes needed to be repeated multiple times, and VIPER® sometimes started the wrong maneuver in response to a command. The maneuver that VIPER® started did not necessarily have a name that was verbally similar to the intended maneuver, indicating that it was a programming problem rather than a limitation in the current capability of voice recognition software. Other than voice commands, the IPs primarily found the updated system easy to use.

These two respondents were involved with the VIPER® evaluation, and were therefore more familiar with VIPER® than other IPs. However, the system was not received well among other instructors at NAS Corpus Christi. Some instructors either had no knowledge of VIPER®, showed no interest in learning about or using VIPER® based on initial perceptions from others, or had a negative impression of the system with first use because of crashing issues. This indicates that the issues associated with the initial development and launch of the VIPER® program made it difficult to obtain buy-in from instructors which in turn made it difficult for students to want to engage with the system.

Meanwhile, based on documented reasons for halting participation, many SNAs either through their own or others’ frustration with the system or through their preconceived beliefs about it, did not expect VIPER® to be useful and therefore did not want to use it. The current evaluation of VIPER® suggests these SNAs’ reported beliefs were unfounded: as mentioned above, VIPER® was associated with performance improvements for SNAs who took the time to use it, as shown both through actual grade comparisons and through the comments of the IPs.

6.4 Recommendations

Based on IP and SNA feedback, the following changes are recommended for future iterations of VIPER® software around three primary areas:

Voice Recognition

- Address incorrect responses to voice commands.
- Until voice recognition technology becomes more reliable, explore possible alternatives to voice commands that do not require using a mouse to navigate through a drop-down menu.

Maneuver Development

- Conduct a more extensive review of VIPER® maneuvers, with a larger number of IPs, to ensure that VIPER® contains an accurate model of the maneuvers and can provide accurate feedback. This may be especially useful for maneuvers that require flexibility in grading.
- VIPER® may need to be limited to maneuvers that require little or only moderate flexibility in grading. If VIPER® does not recognize a non-standard but acceptable way of completing a maneuver, it may incorrectly teach users not to follow good practices. As an example, landing pattern may need to be removed from VIPER®'s list of available maneuvers.

Feedback Delivery

- Where possible, add theoretical information (i.e., why a maneuver should be performed a particular way) into VIPER®'s demonstrations and verbal feedback.
- Present after action feedback in a more useful format, change the grading scale to show percentage of the maneuver correctly completed.

- Allow VIPER® to record and show feedback for maneuvers that were partially completed to enable intentional early exits from maneuvers and to reduce frustration with occasional software crashes.

The following recommendations are made for CNATRA to consider if implementing VIPER® or other ITSs:

- Provide VIPER® on ITDs or similar low-cost simulators as a form of pre-Primary self-practice for SNAs who have completed or during NIFE. This will allow VIPER® to serve as a preliminary, guided introduction to flying in the aircraft.

- Prior to opening VIPER® to SNA usage, give SNAs a brief introduction that includes VIPER®'s potential value (e.g., its ability to help build a sight picture and even improve flight skills on some maneuvers, as well as build correct scan pattern). A similar introduction may also improve VIPER®'s reception among instructors, although the research team recognizes they may not have time to attend an introductory session.

- Improve the accuracy of the flight models. This will make all devices utilizing the models better and specifically for VIPER®, this should help expand the number of maneuvers SNAs could practice appropriately (e.g., aerobatic maneuvers).

- Ensure systems using VIPER® are housed in a quiet enough environment to mitigate voice recognition issues.

Based on research and lessons learned from this study, the team recommends the following considerations when developing and implementing a virtual instructor into training.

- Determine the main objectives of the training that will utilize an ITS, the skills that need to be acquired and that the ITS will address, and the level of competency needed in order to meet the objectives.

- Include experts as early and often in development of the system as possible to provide parameters for acceptable performance and the level of process feedback commonly provided during training with a live instructor. This will enable developers to incorporate expert feedback into appropriate mental models for the system to emulate at the level of detail best suited to the task being trained and the trainee's level of performance on the task.

- Identify the most beneficial elements for the after action review so that feedback is appropriate for the tasks being performed and the level they're being performed at. This may help to avoid detrimental cognitive loading during practice (Billings, 2012).

7. Conclusions

In summary, this evaluation provides empirical evidence of the benefits that VR and ITSs may provide to aviation training. Results document the benefits provided by the introductions of demonstration, deliberate practice, and feedback and provide a basis for considering ITS as an option to help offset limitations in instructor availability when SNAs are preparing for flight events. The researchers' hypotheses were largely supported, as expected based on the expertise and learning science literature, and increases to opportunities for demonstration, deliberate practice, and feedback examined in the study. The results show ITSs, in this case VIPER®, have shown promise in providing training benefits beyond VR practice without feedback, but additional development and study is still needed to fully understand what type of return on investment can be accomplished across aviation training. As a first step towards the acquisition of AI flight training, this evaluation indicates that a supplemental AI instructor does have the potential to improve pilot performance.

8. References

Billings, D. R. (2012). Efficacy of adaptive feedback strategies in simulation-based training. *Military Psychology, 24*, 114-133.

Chief of Naval Air Training (2022). *Mission: Training combat quality aviation professionals.* https://www.cnatra.navy.mil/mission.asp

Decker, P. J. & Nathan, B. R. (1985). *Behavior modeling training: Principles and applications*. New York: Praeger.

Discovery Machine, Inc. (2019). *Virtual Instructor Pilot Exercise Referee (VIPER) For CNATRA.* Topic Number SB052-009 Proposal Number N2-6914.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine, 15,* 988-994.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49,* 725-747.

Ericsson, K.A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363-406.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64,* 349-371.

Kluger, A. N., & DeNisi, A. S. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254-284.

Kozlowski, S. W. J., Bell, B. S., & Mullins, M. E. (2000). *Guiding the development of deployable shipboard training systems: Enhancing skill acquisition, adaptability, and effectiveness* (Contract N61339-96-K-0005). East Lansing, MI: Michigan State University, Department of Psychology.

Landsberg, C. R., Van Buskirk, W. L., & Astwood, R. S. (2010). Does feedback type matter? Investigating the effectiveness of feedback content on performance outcomes. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting- 2010* (pp. 2304-2308).

Lewis, J., & Livingston, J. (2018). Pilot Training Next: Breaking institutional paradigms using student-centered multimodal learning. *Proceedings of the 40th Annual Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Orlando, FL.

McCoy-Fisher, C., Mishler, A., Bush, D., Severe-Valsaint, G., Natali, M., & Riner, B. (2020). *Student naval aviation extended reality device capability evaluation* (Report No. NAWCTSD-TR-2019-001). Fort Belvoir, VA: Defense Technical Information Center.

Rosen, M. A., Salas, E., Pavlas, D., Jensen, R., Fu, D., Lampton, D. (2010). Demonstration-Based Training: A Review of Instructional Features. *Human Factors, 52 (5)*, 596-609.

Salas, E., Rosen, M. A., Pavlas, D., Jensen, R., Fu, D., Ramachandran, S., Hinkelman, E., & Lampton, D. R. (2009). *Understanding Demonstration-based Training: A Definition, Conceptual Framework, and Some Initial Guidelines* (Technical Report No. 1261). Retrieved from Defense Technical Information Center website: https://apps.dtic.mil/sti/pdfs/ADA509390.pdf

Science Applications International Corporation (SAIC). (2018). *Pilot Training Next Iteration 1: Lessons Learned* (Unpublished Technical Report).

Taylor, P. J., Russ-Eft, D.F., & Chan, D.W.L. (2005). A Meta-analytic Review of Behavior Modeling Training. *Journal of Applied Psychology, 90 (4)*, 692-709.

United States Government Accountability Office. (2018). *Military Personnel: DOD needs to reevaluate fighter pilot workforce requirements* (Report No. GAO-18-113). Washington, DC: U.S. Government Accountability Office.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*, 197-221.

VanLehn, K. (2016). Regulative loops, step loops and task loops. *International Journal of Artificial Intelligence in Education, 26*, 107-112.

9. Appendices

9.1. Appendix 1: Event Raw Score Comparisons

Table 3. Mann-Whitney *U* tests on ERS

Event	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
All Contact + Instruments Flights	**Archival vs**	**1.13 (0.05)**	**836**	**917,589.5^{+}**	**.038**	**0.42**
	Free VR	**1.16 (0.07)**	**3014**			
	Archival vs	**1.13 (0.05)**	**836**	**10,032.5^{+}**	**.077**	**0.96**
	Assigned VR	**1.18 (0.08)**	**64**			
	Archival vs	**1.13 (0.05)**	**836**	**4711^{+}**	**.101**	**1.90**
	VIPER®	**1.23 (0.08)**	**52**			
	Free VR vs	**1.16 (0.07)**	**3014**	**60,082.5^{+}**	**.009**	**0.33**
	Assigned VR	**1.18 (0.08)**	**64**			
	Free VR vs	**1.16 (0.07)**	**3014**	**31,326^{+}**	**.018**	**1.05**
	VIPER®	**1.23 (0.08)**	**52**			
	Assigned VR vs	**1.18 (0.08)**	**64**	**1099****	**.086**	**0.60**
	VIPER®	**1.23 (0.08)**	**52**			
All Contact Flights	**Archival vs**	**1.15 (0.06)**	**836**	**933,542.5^{+}**	**.034**	**0.41**
	Free VR	**1.18 (0.07)**	**3014**			
	Archival vs	**1.15 (0.06)**	**836**	**11,966.5^{+}**	**.061**	**0.85**
	Assigned VR	**1.20 (0.09)**	**64**			
	Archival vs	**1.15 (0.06)**	**836**	**6076^{+}**	**.086**	**1.49**
	VIPER®	**1.24 (0.08)**	**52**			
	Free VR vs	**1.18 (0.07)**	**3014**	**65,487.5^{+}**	**.006**	**0.34**
	Assigned VR	**1.20 (0.09)**	**64**			
	Free VR vs	**1.18 (0.07)**	**3014**	**36,876^{+}**	**.014**	**0.88**
	VIPER®	**1.24 (0.08)**	**52**			
	Assigned VR vs	**1.20 (0.09)**	**64**	**1264***	**.043**	**0.49**
	VIPER®	**1.24 (0.08)**	**52**			
All Instruments Flights	**Archival vs**	**1.10 (0.03)**	**836**	**817,064^{+}**	**.015**	**0.29**
	Free VR	**1.11 (0.04)**	**2506**			
	Archival vs	**1.10 (0.03)**	**836**	**8,514^{+}**	**.086**	**1.57**
	Assigned VR	**1.16 (0.05)**	**64**			
	Archival vs	**1.10 (0.03)**	**836**	**4576^{+}**	**.057**	**2.08**
	VIPER®	**1.18 (0.08)**	**36**			
	Free VR vs	**1.11 (0.04)**	**2506**	**38,414^{+}**	**.018**	**0.99**
	Assigned VR	**1.16 (0.05)**	**64**			
	Free VR vs	**1.11 (0.04)**	**2506**	**19,473^{+}**	**.014**	**1.45**
	VIPER®	**1.18 (0.08)**	**36**			
	Assigned VR vs	1.16 (0.05)	64	919	.022	0.34
	VIPER®	1.18 (0.08)	36			
C4101	**Archival vs**	**1.18 (0.13)**	**836**	**892,108.5^{+}**	**.043**	**0.51**
	Free VR	**1.25 (0.14)**	**3010**			
	Archival vs	**1.18 (0.13)**	**836**	**11,251.5^{+}**	**.067**	**1.09**
	Assigned VR	**1.32 (0.15)**	**64**			
	Archival vs	**1.18 (0.13)**	**836**	**9487.5^{+}**	**.052**	**1.15**
	VIPER®	**1.33 (0.14)**	**52**			
	Free VR vs	**1.25 (0.14)**	**3010**	**63,925^{+}**	**.007**	**0.52**
	Assigned VR	**1.32 (0.15)**	**64**			

Event	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
	Free VR vs VIPER®	**1.25 (0.14) 1.33 (0.14)**	**3010 52**	**54,189$^+$**	**.005**	**0.56**
	Assigned VR vs VIPER®	1.32 (0.15) 1.33 (0.14)	64 52	1620.5	<.001	0.04
C4102	**Archival vs Free VR**	**1.27 (0.15) 1.33 (0.15)**	**832 2977**	**943,155.5$^+$**	**.029**	**0.41**
	Archival vs Assigned VR	**1.27 (0.15) 1.42 (0.12)**	**832 63**	**11,334.5$^+$**	**.063**	**1.00**
	Archival vs VIPER®	**1.27 (0.15) 1.45 (0.16)**	**832 52**	**8557.5$^+$**	**.061**	**1.20**
	Free VR vs Assigned VR	**1.33 (0.15) 1.42 (0.12)**	**2977 63**	**60,430$^+$**	**.008**	**0.57**
	Free VR vs VIPER®	**1.33 (0.15) 1.45 (0.16)**	**2977 52**	**45,278.5$^+$**	**.009**	**0.77**
	Assigned VR vs VIPER®	1.42 (0.12) 1.45 (0.16)	63 52	1434.5	.011	0.23
C4103	**Archival vs Free VR**	**1.33 (0.16) 1.38 (0.16)**	**830 2957**	**979,415.5$^+$**	**.021**	**0.34**
	Archival vs Assigned VR	**1.33 (0.16) 1.47 (0.15)**	**830 63**	**12,951$^+$**	**.050**	**0.91**
	Archival vs VIPER®	**1.33 (0.16) 1.51 (0.18)**	**830 52**	**9090$^+$**	**.056**	**1.13**
	Free VR vs Assigned VR	**1.38 (0.16) 1.47 (0.15)**	**2957 63**	**62,458.5$^+$**	**.007**	**0.55**
	Free VR vs VIPER®	**1.38 (0.16) 1.51 (0.18)**	**2957 52**	**43,762$^+$**	**.009**	**0.78**
	Assigned VR vs VIPER®	1.47 (0.15) 1.51 (0.18)	63 52	1421.5	.013	0.23
C4104	**Archival vs Free VR**	**1.39 (0.17) 1.44 (0.17)**	**823 2929**	**998,020.5$^+$**	**.015**	**0.28**
	Archival vs Assigned VR	**1.39 (0.17) 1.52 (0.15)**	**823 63**	**14,393.5$^+$**	**.039**	**0.78**
	Archival vs VIPER®	**1.39 (0.17) 1.55 (0.15)**	**823 52**	**9612$^+$**	**.051**	**0.97**
	Free VR vs Assigned VR	**1.44 (0.17) 1.52 (0.15)**	**2929 63**	**66,227$^+$**	**.005**	**0.50**
	Free VR vs VIPER®	**1.44 (0.17) 1.55 (0.15)**	**2929 52**	**44,560$^+$**	**.009**	**0.69**
	Assigned VR vs VIPER®	1.52 (0.15) 1.55 (0.15)	63 52	1397.5	.016	0.22
C4201	**Archival vs Free VR**	**1.17 (0.11) 1.21 (0.11)**	**823 2892**	**980,049.5$^+$**	**.016**	**0.30**
	Archival vs Assigned VR	**1.17 (0.11) 1.27 (0.10)**	**823 63**	**13,399.5$^+$**	**.046**	**0.88**
	Archival vs VIPER®	**1.17 (0.11) 1.27 (0.09)**	**823 52**	**10,552$^+$**	**.043**	**0.88**
	Free VR vs Assigned VR	**1.21 (0.11) 1.27 (0.10)**	**2892 63**	**61,048$^+$**	**.007**	**0.57**
	Free VR vs VIPER®	**1.21 (0.11) 1.27 (0.09)**	**2892 52**	**49,750$^+$**	**.006**	**0.56**
	Assigned VR vs VIPER®	1.27 (0.10) 1.27 (0.09)	63 52	1600	<.001	-0.01

Event	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
C4202	**Archival vs Free VR**	**1.20 (0.11) 1.23 (0.12)**	**820 2872**	**1,014,961.5$^+$**	**.010**	**0.23**
	Archival vs Assigned VR	**1.20 (0.11) 1.30 (0.11)**	**820 63**	**13,359.5$^+$**	**.046**	**0.89**
	Archival vs VIPER®	**1.20 (0.11) 1.32 (0.08)**	**820 52**	**8599.5$^+$**	**.060**	**1.07**
	Free VR vs Assigned VR	**1.23 (0.12) 1.30 (0.11)**	**2872 63**	**57,764.5$^+$**	**.008**	**0.64**
	Free VR vs VIPER®	**1.23 (0.12) 1.32 (0.08)**	**2872 52**	**39,385$^+$**	**.012**	**0.81**
	Assigned VR vs VIPER®	1.30 (0.11) 1.32 (0.08)	63 52	1503	.005	0.20
C4203	**Archival vs Free VR**	**1.22 (0.12) 1.25 (0.12)**	**818 2836**	**977,980$^+$**	**.013**	**0.27**
	Archival vs Assigned VR	**1.22 (0.12) 1.34 (0.10)**	**818 63**	**11,708$^+$**	**.059**	**1.00**
	Archival vs VIPER®	**1.22 (0.12) 1.34 (0.08)**	**818 51**	**8994.5$^+$**	**.054**	**1.00**
	Free VR vs Assigned VR	**1.25 (0.12) 1.34 (0.10)**	**2836 63**	**51,887.5$^+$**	**.011**	**0.73**
	Free VR vs VIPER®	**1.25 (0.12) 1.34 (0.08)**	**2836 51**	**40,577$^+$**	**.010**	**0.73**
	Assigned VR vs VIPER®	1.34 (0.10) 1.34 (0.08)	63 51	1567.5	<.001	-0.01
C4204	**Archival vs Free VR**	**1.25 (0.11) 1.27 (0.12)**	**814 2825**	**1,031,073$^+$**	**.006**	**0.17**
	Archival vs Assigned VR	**1.25 (0.11) 1.33 (0.10)**	**814 61**	**14,975.5$^+$**	**.031**	**0.71**
	Archival vs VIPER®	**1.25 (0.11) 1.33 (0.10)**	**814 49**	**12,164.5$^+$**	**.024**	**0.70**
	Free VR vs Assigned VR	**1.27 (0.12) 1.33 (0.10)**	**2825 61**	**60,927$^+$**	**.005**	**0.52**
	Free VR vs VIPER®	**1.27 (0.12) 1.33 (0.10)**	**2825 49**	**49,371$^+$**	**.004**	**0.51**
	Assigned VR vs VIPER®	1.33 (0.10) 1.33 (0.10)	61 49	1478.5	<.001	-0.01
C4301	**Archival vs Free VR**	**1.07 (0.05) 1.08 (0.05)**	**813 2809**	**986,453.5$^+$**	**.010**	**0.25**
	Archival vs Assigned VR	**1.07 (0.05) 1.09 (0.05)**	**813 61**	**19,219.5****	**.010**	**0.42**
	Archival vs VIPER®	**1.07 (0.05) 1.10 (0.04)**	**813 46**	**12,196.5$^+$**	**.018**	**0.60**
	Free VR vs Assigned VR	1.08 (0.05) 1.09 (0.05)	2809 61	77,577.5	.001	0.18
	Free VR vs VIPER®	**1.08 (0.05) 1.10 (0.04)**	**2809 46**	**51,051.5***	**.002**	**0.35**
	Assigned VR vs VIPER®	1.09 (0.05) 1.10 (0.04)	61 46	1264	.007	0.18
C4302	**Archival vs Free VR**	**1.08 (0.05) 1.09 (0.05)**	**811 2787**	**1,005,997.5$^+$**	**.006**	**0.21**
	Archival vs Assigned VR	**1.08 (0.05) 1.10 (0.05)**	**811 61**	**18,655.5****	**.012**	**0.45**

Event	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
	Archival vs VIPER®	**1.08 (0.05) 1.11 (0.03)**	**811 46**	**12,652.5^{+}**	**.016**	**0.55**
	Free VR vs Assigned VR	1.09 (0.05) 1.10 (0.05)	2787 61	72,592°	.001	0.25
	Free VR vs VIPER®	**1.09 (0.05) 1.11 (0.03)**	**2787 46**	**51,311***	**.002**	**0.34**
	Assigned VR vs VIPER®	1.10 (0.05) 1.11 (0.03)	61 46	1380	<.001	0.10
C4303	**Archival vs Free VR**	**1.09 (0.05) 1.10 (0.05)**	**809 2741**	**971,009^{+}**	**.008**	**0.23**
	Archival vs Assigned VR	1.09 (0.05) 1.10 (0.04)	809 53	19,142.5	.002	0.21
	Archival vs VIPER®	**1.09 (0.05) 1.12 (0.04)**	**809 42**	**11,099.5^{+}**	**.017**	**0.63**
	Free VR vs Assigned VR	1.10 (0.05) 1.10 (0.04)	2741. 53	70,645	<.001	-0.02
	Free VR vs VIPER®	**1.10 (0.05) 1.12 (0.04)**	**2741 42**	**44,674.5***	**.002**	**0.42**
	Assigned VR vs VIPER®	**1.10 (0.04) 1.12 (0.04)**	**53 42**	**807***	**.056**	**0.53**
C4304	**Archival vs Free VR**	**1.10 (0.04) 1.11 (0.05)**	**809 2736**	**977,585.5^{+}**	**.007**	**0.20**
	Archival vs Assigned VR	1.10 (0.04) 1.11 (0.04)	809 50	18,296.5	.001	0.16
	Archival vs VIPER®	**1.10 (0.04) 1.13 (0.04)**	**809 39**	**8948.5^{+}**	**.025**	**0.80**
	Free VR vs Assigned VR	1.11 (0.05) 1.11 (0.04)	2736 50	65,922.5	<.001	-0.05
	Free VR vs VIPER®	**1.11 (0.05) 1.13 (0.04)**	**2736 39**	**36,553^{+}**	**.004**	**0.54**
	Assigned VR vs VIPER®	**1.11 (0.04) 1.13 (0.04)**	**50 39**	**607****	**.105**	**0.75**
C4490	**Archival vs Free VR**	**1.09 (0.04) 1.10 (0.05)**	**807 2762**	**997,291^{+}**	**.006**	**0.17**
	Archival vs Assigned VR	**1.09 (0.04) 1.10 (0.04)**	**807 63**	**21,448.5***	**.005**	**0.29**
	Archival vs VIPER®	**1.09 (0.04) 1.12 (0.03)**	**807 45**	**13,120****	**.012**	**0.47**
	Free VR vs Assigned VR	1.10 (0.05) 1.10 (0.04)	2762 63	82,861	<.001	0.10
	Free VR vs VIPER®	1.10 (0.05) 1.12 (0.03)	2762 45	51,909°	.001	0.26
	Assigned VR vs VIPER®	1.10 (0.04) 1.12 (0.03)	63 45	1243	.011	0.21
C4601	**Archival vs Free VR**	**1.04 (0.03) 1.05 (0.03)**	**799 2715**	**973,426.5^{+}**	**.006**	**0.19**
	Archival vs Assigned VR	**1.04 (0.03) 1.06 (0.03)**	**799 63**	**18,133^{+}**	**.016**	**0.49**
	Archival vs VIPER®	**1.04 (0.03) 1.06 (0.03)**	**799 42**	**11,863.5****	**.012**	**0.46**
	Free VR vs Assigned VR	**1.05 (0.03) 1.06 (0.03)**	**2715 63**	**70,888.5***	**.002**	**0.28**

Event	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
	Free VR vs VIPER®	**1.05 (0.03) 1.06 (0.03)**	**2715 42**	**46,300.5***	**.002**	**0.25**
	Assigned VR vs VIPER®	1.06 (0.03) 1.06 (0.03)	63 42	1286	<.001	-0.03
C4602	**Archival vs Free VR**§	**1.06 (0.04) 1.06 (0.04)**	**797 2698**	**1,014,094.5***	**.002**	**0.12**
	Archival vs Assigned VR	**1.06 (0.04) 1.07 (0.04)**	**797 62**	**21,010***	**.005**	**0.30**
	Archival vs VIPER®	1.06 (0.04) 1.06 (0.04)	797 41	15,270.5	<.001	0.12
	Free VR vs Assigned VR	1.06 (0.04) 1.07 (0.04)	2698 62	75,723	.001	0.18
	Free VR vs VIPER®	1.06 (0.04) 1.06 (0.04)	2698 41	54,643	<.001	0.01
	Assigned VR vs VIPER®	1.07 (0.04) 1.06 (0.04)	62 41	1162	.005	-0.17
C4603	Archival vs Free VR	1.06 (0.04) 1.07 (0.04)	797 2679	1,024,679.5°	.001	0.10
	Archival vs Assigned VR	**1.06 (0.04) 1.07 (0.04)**	**797 61**	**20,429.5***	**.005**	**0.32**
	Archival vs VIPER®	1.06 (0.04) 1.07 (0.04)	797 41	13,557.5°	.004	0.31
	Free VR vs Assigned VR	1.07 (0.04) 1.07 (0.04)	2679 61	71,811.5	.001	0.22
	Free VR vs VIPER®	1.07 (0.04) 1.07 (0.04)	2679 41	47,589.5	.001	0.21
	Assigned VR vs VIPER®	1.07 (0.04) 1.07 (0.04)	61 41	1228.5	<.001	-0.01
C4604	Archival vs Free VR	1.06 (0.04) 1.07 (0.04)	794 2677	1,035,419	<.001	0.05
	Archival vs Assigned VR	**1.06 (0.04) 1.08 (0.03)**	**794 63**	**19,976****	**.008**	**0.32**
	Archival vs VIPER®	**1.06 (0.04) 1.08 (0.03)**	**794 41**	**12,176.5****	**.009**	**0.41**
	Free VR vs Assigned VR	**1.07 (0.04) 1.08 (0.03)**	**2677 63**	**69,626***	**.002**	**0.27**
	Free VR vs VIPER®	**1.07 (0.04) 1.08 (0.03)**	**2677 41**	**42,325***	**.002**	**0.36**
	Assigned VR vs VIPER®	1.08 (0.03) 1.08 (0.03)	63 41	1229.5	.002	0.10
C4790	**Archival vs Free VR**§	**1.06 (0.03) 1.06 (0.03)**	**794 2671**	**966,607**$^{+}$	**.004**	**0.14**
	Archival vs Assigned VR	**1.06 (0.03) 1.08 (0.03)**	**794 63**	**15,673.5**$^{+}$	**.028**	**0.66**
	Archival vs VIPER®	**1.06 (0.03) 1.07 (0.03)**	**794 40**	**10,280.5**$^{+}$	**.017**	**0.61**
	Free VR vs Assigned VR	**1.06 (0.03) 1.08 (0.03)**	**2671 63**	**59,699.5**$^{+}$	**.006**	**0.48**
	Free VR vs VIPER®	**1.06 (0.03) 1.07 (0.03)**	**2671 40**	**39,735****	**.003**	**0.43**
	Assigned VR vs VIPER®	1.08 (0.03) 1.07 (0.03)	63 40	1210	.001	-0.05

Event	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
C4901	**Archival vs Free VR**	**1.10 (0.04)** **1.11 (0.05)**	**788** **2527**	**875,183⁺**	**.008**	**0.14**
	Archival vs Assigned VR	**1.10 (0.04)** **1.12 (0.05)**	**788** **62**	**17,264⁺**	**.018**	**0.55**
	Archival vs VIPER®	**1.10 (0.04)** **1.13 (0.03)**	**788** **37**	**9077⁺**	**.018**	**0.62**
	Free VR vs Assigned VR	**1.11 (0.05)** **1.12 (0.05)**	**2527** **62**	**65,317.5***	**.002**	**0.35**
	Free VR vs VIPER®	**1.11 (0.05)** **1.13 (0.03)**	**2527** **37**	**35,844.5***	**.002**	**0.40**
	Assigned VR vs VIPER®	1.12 (0.05) 1.13 (0.03)	62 37	1085.5	.002	0.05
I4101	**Archival vs Free VR**	**1.18 (0.11)** **1.20 (0.10)**	**781** **2504**	**877,405.5⁺**	**.006**	**0.18**
	Archival vs Assigned VR	**1.18 (0.11)** **1.24 (0.09)**	**781** **62**	**15,902⁺**	**.024**	**0.60**
	Archival vs VIPER®	**1.18 (0.11)** **1.26 (0.09)**	**781** **36**	**7802.5⁺**	**.025**	**0.76**
	Free VR vs Assigned VR	**1.20 (0.10)** **1.24 (0.09)**	**2504** **62**	**58,401⁺**	**.004**	**0.43**
	Free VR vs VIPER®	**1.20 (0.10)** **1.26 (0.09)**	**2504** **36**	**28,827.5⁺**	**.005**	**0.59**
	Assigned VR vs VIPER®	1.24 (0.09) 1.26 (0.09)	62 36	957.5	.014	0.19
I4102	**Archival vs Free VR**	**1.21 (0.10)** **1.22 (0.10)**	**781** **2493**	**906,150****	**.003**	**0.12**
	Archival vs Assigned VR	**1.21 (0.10)** **1.27 (0.07)**	**781** **59**	**15,422.5⁺**	**.021**	**0.61**
	Archival vs VIPER®	**1.21 (0.10)** **1.25 (0.09)**	**781** **36**	**10,303****	**.009**	**0.44**
	Free VR vs Assigned VR	**1.22 (0.10)** **1.27 (0.07)**	**2493** **59**	**54,066⁺**	**.005**	**0.49**
	Free VR vs VIPER®	**1.22 (0.10)** **1.25 (0.09)**	**2493** **36**	**35,906***	**.002**	**0.32**
	Assigned VR vs VIPER®	1.27 (0.07) 1.25 (0.09)	59 36	989	.003	-0.21
I4103	**Archival vs Free VR**	**1.23 (0.10)** **1.24 (0.10)**	**779** **2477**	**909,119.5***	**.002**	**0.11**
	Archival vs Assigned VR	**1.23 (0.10)** **1.24 (0.10)**	**779** **58**	**14,693.5⁺**	**.024**	**0.57**
	Archival vs VIPER®	1.23 (0.10) 1.26 (0.09)	779 34	10,904.5°	.004	0.31
	Free VR vs Assigned VR	**1.24 (0.10)** **1.29 (0.09)**	**2477** **58**	**50,676⁺**	**.006**	**0.47**
	Free VR vs VIPER®	1.24 (0.10) 1.26 (0.09)	2477 34	36,923.5	.001	0.21
	Assigned VR vs VIPER®	1.29 (0.09) 1.26 (0.09)	58 34	792.5	.027	-0.29
I4104	**Archival vs Free VR**	**1.25 (0.10)** **1.26 (0.10)**	**779** **2462**	**886,219****	**.003**	**0.13**
	Archival vs Assigned VR	**1.25 (0.10)** **1.29 (0.08)**	**779** **60**	**15,872.5⁺**	**.021**	**0.51**

Event	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
	Archival vs VIPER®	1.25 (0.10) 1.27 (0.10)	779 31	9980.5	.003	0.29
	Free VR vs Assigned VR	**1.26 (0.10) 1.29 (0.08)**	**2462 60**	**54,849^{+}**	**.005**	**0.37**
	Free VR vs VIPER®	1.26 (0.10) 1.27 (0.10)	2462 31	34,051	<.001	0.16
	Assigned VR vs VIPER®	1.29 (0.08) 1.27 (0.10)	60 31	801	.013	-0.24
I4201	**Archival vs Free VR[§]**	**1.08 (0.05) 1.08 (0.04)**	**779 2443**	**82,792^{+}**	**.006**	**0.18**
	Archival vs Assigned VR	**1.08 (0.05) 1.10 (0.04)**	**779 54**	**13,825^{+}**	**.021**	**0.55**
	Archival vs VIPER®	**1.08 (0.05) 1.11 (0.05)**	**779 27**	**7,330****	**.009**	**0.64**
	Free VR vs Assigned VR	**1.08 (0.04) 1.10 (0.04)**	**2443 54**	**50,164.5****	**.004**	**0.38**
	Free VR vs VIPER®	1.08 (0.04) 1.11 (0.05)	2443 27	25,914^{o}	.001	0.48
	Assigned VR vs VIPER®	1.10 (0.04) 1.11 (0.05)	54 27	722.5	<.001	0.10
I4202	**Archival vs Free VR**	**1.08 (0.04) 1.09 (0.04)**	**777 2396**	**877,406.5***	**.002**	**0.10**
	Archival vs Assigned VR	**1.08 (0.04) 1.10 (0.05)**	**777 50**	**15,542***	**.007**	**0.50**
	Archival vs VIPER®	1.08 (0.04) 1.10 (0.06)	777 25	7678^{o}	.004	0.47
	Free VR vs Assigned VR	1.09 (0.04) 1.10 (0.05)	2396 50	50,761^{o}	.001	0.35
	Free VR vs VIPER®	1.09 (0.04) 1.10 (0.06)	2396 25	25,683	.001	0.33
	Assigned VR vs VIPER®	1.10 (0.05) 1.10 (0.06)	50 25	618.5	<.001	-0.03
I4203	**Archival vs Free VR**	**1.08 (0.04) 1.09 (0.04)**	**773 2323**	**809,946^{+}**	**.005**	**0.16**
	Archival vs Assigned VR	**1.08 (0.04) 1.10 (0.04)**	**773 34**	**10,102.5***	**.006**	**0.47**
	Archival vs VIPER®	1.08 (0.04) 1.09 (0.05)	773 19	6813	<.001	0.20
	Free VR vs Assigned VR	1.09 (0.04) 1.10 (0.04)	2323 34	34,448	.001	0.25
	Free VR vs VIPER®	1.09 (0.04) 1.09 (0.05)	2323 19	21,713.5	<.001	0.02
	Assigned VR vs VIPER®	1.10 (0.04) 1.09 (0.05)	34 19	278	.013	-0.24
I4204	**Archival vs Free VR[§]**	**1.09 (0.04) 1.09 (0.04)**	**771 2352**	**812,781.5^{+}**	**.006**	**0.16**
	Archival vs Assigned VR	1.09 (0.04) 1.10 (0.03)	771 35	11,734	.002	0.21
	Archival vs VIPER®	1.09 (0.04) 1.10 (0.05)	771 17	5076.5	.003	0.41
	Free VR vs Assigned VR	1.09 (0.04) 1.10 (0.03)	2352 35	40,620.5	<.001	0.02

Event	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
	Free VR vs VIPER®	1.09 (0.04) 1.10 (0.05)	2352 17	17,432.5	<.001	0.20
	Assigned VR vs VIPER®	1.10 (0.03) 1.10 (0.05)	35 17	262.5	.009	0.19
I4301	**Archival vs Free VR[§]**	**1.03 (0.02) 1.03 (0.03)**	**776 2427**	**892,379.5***	**.002**	**0.10**
I4301	**Archival vs Assigned VR**	**1.03 (0.02) 1.04 (0.03)**	**776 52**	**15,213.5****	**.011**	**0.45**
I4301	Archival vs VIPER®	1.03 (0.02) 1.04 (0.04)	776 23	8680	<.001	0.18
I4301	**Free VR vs Assigned VR**	**1.03 (0.03) 1.04 (0.03)**	**2427 52**	**51,449***	**.002**	**0.30**
I4301	Free VR vs VIPER®	1.03 (0.03) 1.04 (0.04)	2427 23	27,425.5	<.001	0.06
I4301	Assigned VR vs VIPER®	1.04 (0.03) 1.04 (0.04)	52 23	497	.018	-0.22
I4302	**Archival vs Free VR**	**1.03 (0.03) 1.04 (0.03)**	**775 2422**	**843,345[+]**	**.006**	**0.17**
I4302	**Archival vs Assigned VR**	**1.03 (0.03) 1.04 (0.02)**	**775 51**	**15,663***	**.007**	**0.29**
I4302	**Archival vs VIPER®**	**1.03 (0.03) 1.05 (0.03)**	**775 23**	**5517.5****	**.012**	**0.68**
I4302	Free VR vs Assigned VR	1.04 (0.03) 1.04 (0.02)	2422 51	55,507.5	.001	0.07
I4302	**Free VR vs VIPER®**	**1.04 (0.03) 1.05 (0.03)**	**2422 23**	**20,135.5***	**.002**	**0.40**
I4302	Assigned VR vs VIPER®	1.04 (0.02) 1.05 (0.03)	51 23	455	.032	0.38
I4303	**Archival vs Free VR**	**1.03 (0.03) 1.04 (0.03)**	**775 2262**	**811,724****	**.003**	**0.11**
I4303	Archival vs Assigned VR	1.03 (0.03) 1.03 (0.02)	775 27	10,348	<.001	-0.06
I4303	Archival vs VIPER®	1.03 (0.03) 1.03 (0.02)	775 11	3878	<.001	-0.16
I4303	Free VR vs Assigned VR	1.04 (0.03) 1.03 (0.02)	2262 27	28,794.5	<.001	-0.17
I4303	Free VR vs VIPER®	1.04 (0.03) 1.03 (0.02)	2262 11	10,298	<.001	-0.26
I4303	Assigned VR vs VIPER®	1.03 (0.02) 1.03 (0.02)	27 11	130.5	.009	-0.14
I4304	**Archival vs Free VR**	**1.03 (0.03) 1.04 (0.03)**	**773 2235**	**776,218[+]**	**.006**	**0.17**
I4304	Archival vs Assigned VR	1.03 (0.03) 1.03 (0.02)	773 21	7,614.5	<.001	0.03
I4304	Archival vs VIPER®	1.03 (0.03) 1.03 (0.02)	773 9	3444.5	<.001	-0.01
I4304	Free VR vs Assigned VR	1.04 (0.03) 1.03 (0.02)	2235 21	22,317	<.001	-0.14
I4304	Free VR vs VIPER®	1.04 (0.03) 1.03 (0.02)	2235 9	9,135.5	<.001	-0.17
I4304	Assigned VR vs VIPER®	1.03 (0.02) 1.03 (0.02)	21 9	91	<.001	-0.04

Event	Comparison	M(SD)	n	U	η^2	g
I4305	**Archival vs Free VR**	**1.03 (0.02)** **1.04 (0.03)**	**773** **2310**	**739,672⁺**	**.017**	**0.29**
	Archival vs Assigned VR	1.03 (0.02) 1.04 (0.02)	773 27	8,146,5°	.004	0.31
	Archival vs VIPER®	1.03 (0.02) 1.05 (0.03)	773 15	4468.5	.003	0.46
	Free VR vs Assigned VR	1.04 (0.03) 1.04 (0.02)	2310 27	30,101.5	<.001	-0.03
	Free VR vs VIPER®	1.04 (0.03) 1.05 (0.03)	2310 15	16,293.5	<.001	0.10
	Assigned VR vs VIPER®	1.04 (0.02) 1.05 (0.03)	27 15	191.5	.002	0.13
I4490	**Archival vs Free VR**	**1.04 (0.02)** **1.05 (0.03)**	**774** **2407**	**778,138.5⁺**	**.015**	**0.28**
	Archival vs Assigned VR	**1.04 (0.02)** **1.05 (0.03)**	**774** **52**	**15,066****	**.011**	**0.41**
	Archival vs VIPER®	**1.04 (0.02)** **1.05 (0.03)**	**774** **21**	**5898***	**.006**	**0.61**
	Free VR vs Assigned VR	1.05 (0.03) 1.05 (0.03)	2407 52	58,907.5	<.001	0.04
	Free VR vs VIPER®	1.05 (0.03) 1.05 (0.03)	2407 21	22,303.5	<.001	0.19
	Assigned VR vs VIPER®	1.05 (0.03) 1.05 (0.03)	52 21	506.5	.003	0.17

Note. *M* and *SD* = mean and standard deviation, *n* = number of participants included in the Mann-Whitney *U* test, *U* = Mann-Whitney *U* statistic, η^2 = effect size for Mann-Whitney *U* test (proportion of variation attributable to difference in ranks), *g* = Hedges' *g* (difference between groups in standard deviation units). Significant Mann-Whitney *U* tests are indicated with bold text. * = $p < .05$, ** = $p < .01$, ⁺ = $p < .001$, ° = $p < .1$.

§Although the means rounded to 2 decimal places appear identical in these two conditions, the Free VR group had a higher mean rank than the Archival group, indicating a tendency toward higher ERS in the Free VR group. C4602: Archival Mean Rank = 1671.39, Free VR Mean Rank = 1770.63. C4790: Archival Mean Rank = 1614.89, Free VR Mean Rank = 1768.11. I4201: Archival Mean Rank = 1484.73, Free VR Mean Rank = 1651.92. I4204: Archival Mean Rank = 1440.19, Free VR Mean Rank = 1601.93. I4301: Archival Mean Rank = 1538.47, Free VR Mean Rank = 1622.31.

9.2. Appendix 2: Modifier Code Comparisons

Table 4. Mann-Whitney *U* tests on percentage of event modifiers

Event Modifier	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
Adaptation Sortie (Contact + Instruments Flights)	Archival vs Free VR	0.04% (0.47%) 0.05% (0.66%)	836 3014	1,259,522.5	<.001	0.01
	All others N/A: Assigned VR and VIPER® conditions did not have any Adaptation Sorties					
Adaptation Sortie (Contact Flights)	Archival vs Free VR	0.05% (0.57%) 0.06% (0.80%)	836 3014	1,259,521	<.001	0.02
	All others N/A: Assigned VR and VIPER® conditions did not have any Adaptation Sorties					
Adaptation Sortie (Instruments Flights)	N/A: There were no Adaptation Sorties for Instruments flights in any condition					
Practice Sortie (Contact + Instruments Flights)	Archival vs Free VR	<0.01% (0.10%) <0.01% (0.08%)	836 3014	1,259,181.5	<.001	-0.02
	All others N/A: Assigned VR and VIPER® conditions did not have any Practice Sorties					
Practice Sortie (Contact Flights)	Archival vs Free VR	0.01% (0.18%) <0.01% (0.10%)	836 3014	1,258,763	<.001	-0.04
	All others N/A: Assigned VR and VIPER® conditions did not have any Practice Sorties					
Practice Sortie (Instruments Flights)	N/A: Archival, Assigned VR, and VIPER® did not have any Practice Sorties for Instruments flights					
Warmup Sortie (Contact + Instruments Flights)	Archival vs Free VR	3.43% (4.39%) 3.77% (6.76%)	836 3014	1,227,788.5	<.001	0.05
	Archival vs Assigned VR	3.43% (4.39%) 3.14% (3.55%)	836 64	25,900	<.001	-0.07
	Archival vs VIPER®	**3.43% (4.39%) 2.20% (3.81%)**	**836 52**	**17,265****	**.008**	**-0.28**
	Free VR vs Assigned VR	3.77% (6.76%) 3.14% (3.55%)	3014 64	92,400	<.001	-0.09
	Free VR vs VIPER®	**3.77% (6.76%) 2.20% (3.81%)**	**3014 52**	**64,826.5***	**.002**	**-0.23**
	Assigned VR vs VIPER®	**3.14% (3.55%) 2.20% (3.81%)**	**64 52**	**1317.5***	**.037**	**-0.26**
Warmup Sortie (Contact Flights)	Archival vs Free VR	5.06% (5.92%) 5.21% (7.87%)	836 3014	1,221,332	.001	0.02
	Archival vs Assigned VR	5.06% (5.92%) 4.37% (4.66%)	836 64	26,368	<.001	-0.12
	Archival vs VIPER®	**5.06% (5.92%) 2.18% (3.98%)**	**836 52**	**15,845^{+}**	**.014**	**-0.49**
	Free VR vs Assigned VR	5.21% (7.87%) 4.37% (4.66%)	3014 64	94,867.5	<.001	-0.11
	Free VR vs VIPER®	**5.21% (7.87%) 2.18% (3.98%)**	**3014 52**	**60,160.5****	**.003**	**-0.39**

Event Modifier	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
	Assigned VR vs VIPER®	**4.37% (4.66%) 2.18% (3.98%)**	**64 52**	**1223.5****	**.010**	**-0.50**
Warmup Sortie (Instruments Flights)	**Archival vs Free VR**	**0.70% (2.36%) 0.51% (2.12%)**	**781 2506**	**954,935.5***	**.002**	**-0.09**
	Archival vs Assigned VR	0.70% (2.36%) 0.41% (2.28%)	781 62	22,940	.002	-0.12
	Archival vs VIPER®	0.70% (2.36%) 2.02% (8.49%)	781 36	13,670	<.001	0.45
	Free VR vs Assigned VR	0.51% (2.12%) 0.41% (2.28%)	2506 62	75,457.5	<.001	-0.05
	Free VR vs VIPER®	0.51% (2.12%) 2.02% (8.49%)	2506 36	42,818.5	.001	0.65
	Assigned VR vs VIPER®	0.41% (2.28%) 2.02% (8.49%)	62 36	1030	.024	0.30
Extra Training (Contact + Instruments Flights)	Archival vs Free VR	0.04% (0.34%) 0.04% (0.49%)	836 3014	1,253,767	<.001	-0.01
	Archival vs Assigned VR	0.04% (0.34%) 0.08% (0.66%)	836 64	26,680.5	<.001	0.11
	Archival vs VIPER®	0.04% (0.34%) 0.15% (1.07%)	836 52	21,598.5	<.001	0.26
	Free VR vs Assigned VR	0.04% (0.49%) 0.08% (0.66%)	3014 64	95,731.5	<.001	0.09
	Free VR vs VIPER®	0.04% (0.49%) 0.15% (1.07%)	3014 52	77,497.5	<.001	0.22
	Assigned VR vs VIPER®	0.08% (0.66%) 0.15% (1.07%)	64 52	1657.5	<.001	0.08
Extra Training (Contact Flights)	Archival vs Free VR	0.02% (0.35%) 0.05% (0.58%)	836 3014	1,255,169.5	<.001	0.05
	Archival vs Assigned VR	0.02% (0.35%) 0.08% (0.66%)	836 64	26,430	.002	0.18
	Archival vs VIPER®	0.02% (0.35%) 0.15% (1.07%)	836 52	21,394.5	.003	0.32
	Free VR vs Assigned VR	0.05% (0.58%) 0.08% (0.66%)	3014 64	95,647.5	<.001	0.06
	Free VR vs VIPER®	0.05% (0.58%) 0.15% (1.07%)	3014 52	77,421	<.001	0.17
	Assigned VR vs VIPER®	0.08% (0.66%) 0.15% (1.07%)	64 52	1657.5	<.001	0.08
Extra Training (Instruments Flights)	**Archival vs Free VR**	**0.07% (0.72%) 0.01% (0.25%)**	**781 2506**	**969,740.5^{+}**	**.004**	**-0.16**
	All others N/A: Assigned VR and VIPER® conditions did not have any Extra Training in Instruments events					
Progress Checkride (Contact + Instruments Flights)	Archival vs Free VR	0.21% (0.99%) 0.27% (1.04%)	836 3014	1,238,631.5^{o}	.001	0.06
	Archival vs Assigned VR	0.21% (0.99%) 0.26% (0.93%)	836 64	26,200	<.001	0.05
	Archival vs VIPER®	0.21% (0.99%) 0.37% (1.38%)	836 52	21,280.5	<.001	0.15
	Free VR vs Assigned VR	0.27% (1.04%) 0.26% (0.93%)	3014 64	96,083	<.001	-0.01
	Free VR vs VIPER®	0.27% (1.04%) 0.37% (1.38%)	3014 52	78,002	<.001	0.09

Event Modifier	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
	Assigned VR vs VIPER®	0.26% (0.93%) 0.37% (1.38%)	64 52	1661.5	<.001	0.09
Progress Checkride (Contact Flights)	Archival vs Free VR	0.33% (1.38%) 0.39% (1.44%)	836 3014	1,243,945.5	<.001	0.04
	Archival vs Assigned VR	0.33% (1.38%) 0.42% (1.44%)	836 64	26,220	<.001	0.06
	Archival vs VIPER®	0.33% (1.38%) 0.46% (1.62%)	836 52	21,301.5	<.001	0.09
	Free VR vs Assigned VR	0.39% (1.44%) 0.42% (1.44%)	3014 64	95,741.5	<.001	0.02
	Free VR vs VIPER®	0.39% (1.44%) 0.46% (1.62%)	3014 52	77,759	<.001	0.04
	Assigned VR vs VIPER®	0.42% (1.44%) 0.46% (1.62%)	64 52	1662.5	<.001	0.03
Progress Checkride (Instruments Flights)	N/A: Archival, Assigned VR, and VIPER® groups did not have any Progress Checkrides for Instruments flights					
Elimination Checkride (Contact + Instruments Flights)	Archival vs Free VR	0.36% (3.90%) 0.24% (2.31%)	836 3014	1,257,271	<.001	-0.05
	All others N/A: Assigned VR and VIPER® conditions did not have any elimination checkrides					
Elimination Checkride (Contact Flights)	Archival vs Free VR	0.38% (3.93%) 0.24% (2.33%)	836 3014	1,253,001	<.001	-0.05
	All others N/A: Assigned VR and VIPER® conditions did not have any elimination checkrides					
Elimination Checkride (Instruments Flights)	Archival vs Free VR	0.02% (0.51%) 0.08% (2.07%)	781 2506	973,995^{o}	.001	0.03
	All others N/A: Assigned VR and VIPER® conditions did not have any elimination checkrides					
Repeat (Contact + Instruments Flights)	**Archival vs Free VR**	**3.91% (4.60%) 3.39% (5.42%)**	**836 3014**	**1,117,509**$^{+}$	**.007**	**-0.10**
	Archival vs Assigned VR	**3.91% (4.60%) 0.83% (1.81%)**	**836 64**	**15,348.5**$^{+}$	**.040**	**-0.69**
	Archival vs VIPER®	**3.91% (4.60%) 2.27% (4.03%)**	**836 52**	**16,668.5****	**.010**	**-0.36**
	Free VR vs Assigned VR	**3.39% (5.42%) 0.83% (1.81%)**	**3014 64**	**66,761.5**$^{+}$	**.007**	**-0.48**
	Free VR vs VIPER®	3.39% (5.42%) 2.27% (4.03%)	3014 52	68,607^{o}	.001	-0.21
	Assigned VR vs VIPER®	**0.83% (1.81%) 2.27% (4.03%)**	**64 52**	**1375***	**.037**	**0.48**
Repeat (Contact Flights)	**Archival vs Free VR**	**5.61% (6.49%) 4.63% (6.83%)**	**836 3014**	**1,123,997**$^{+}$	**.007**	**-0.14**
	Archival vs Assigned VR	**5.61% (6.49%) 0.99% (2.64%)**	**836 64**	**14,976**$^{+}$	**.044**	**-0.73**
	Archival vs VIPER®	**5.61% (6.49%) 2.23% (4.28%)**	**836 52**	**15,114.5**$^{+}$	**.017**	**-0.53**
	Free VR vs Assigned VR	**4.63% (6.83%) 0.99% (2.64%)**	**3014 64**	**65,055.5**$^{+}$	**.008**	**-0.54**

Event Modifier	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*
	Free VR vs VIPER®	**4.63% (6.83%) 2.23% (4.28%)**	**3014 52**	**63,179.5****	**.002**	**-0.35**
	Assigned VR vs VIPER®	0.99% (2.64%) 2.23% (4.28%)	64 52	1445°	.026	0.36
Repeat (Instruments Flights)	**Archival vs Free VR**	**1.22% (3.05%) 0.90% (2.88%)**	**781 2506**	**938,123.5****	**.003**	**-0.11**
	Archival vs Assigned VR	**1.22% (3.05%) 0.41% (1.84%)**	**781 62**	**21,832***	**.021**	**-0.27**
	Archival vs VIPER®	1.22% (3.05%) 1.47% (4.81%)	781 36	13,699	<.001	-0.08
	Free VR vs Assigned VR	0.90% (2.88%) 0.41% (1.84%)	2506 62	73,216.5	.001	-0.17
	Free VR vs VIPER®	0.90% (2.88%) 1.47% (4.81%)	2506 36	44,547	<.001	0.19
	Assigned VR vs VIPER®	0.41% (1.84%) 1.47% (4.81%)	62 36	1044	.015	0.33
Unsat (Contact + Instruments Flights)	**Archival vs Free VR**	**1.53% (5.13%) 0.32% (3.51%)**	**836 3014**	**964,725.5⁺**	**.125**	**-0.31**
	Archival vs Assigned VR§	**1.53% (5.13%) 1.56% (12.50%)**	**836 64**	**20,145⁺**	**.021**	**0.01**
	Archival vs VIPER®	N/A: VIPER® condition did not have any unsats				
	Free VR vs Assigned VR	0.32% (3.51%) 1.56% (12.50%)	3014 64	95,181.5	<.001	0.32
	Free VR vs VIPER®	N/A: VIPER® condition did not have any unsats				
	Assigned VR vs VIPER®	N/A: VIPER® condition did not have any unsats				
Unsat (Contact Flights)	**Archival vs Free VR**	**2.04% (5.54%) 0.33% (3.53%)**	**836 3014**	**979,131.5⁺**	**.119**	**-0.42**
	Archival vs Assigned VR	**2.04% (5.54%) 1.56% (12.50%)**	**836 64**	**20,523⁺**	**.020**	**-0.08**
	Archival vs VIPER®	N/A: VIPER® condition did not have any unsats				
	Free VR vs Assigned VR	0.33% (3.53%) 1.56% (12.50%)	3014 64	95,339	<.001	0.31
	Free VR vs VIPER®	N/A: VIPER® condition did not have any unsats				
	Assigned VR vs VIPER®	N/A: VIPER® condition did not have any unsats				
Unsat (Instruments Flights)	**Archival vs Free VR**	**0.24% (1.49%) 0.09% (2.85%)**	**781 2506**	**951,756.5⁺**	**.016**	**-0.06**
	All others N/A: Assigned VR and VIPER® conditions did not have any unsats in Instruments flights					

Note. *M* and *SD* = mean and standard deviation, *n* = number of participants included in the Mann-Whitney *U* test, *U* = Mann-Whitney *U* statistic, η^2 = effect size for Mann-Whitney *U* test (proportion of variation attributable to difference in ranks), *g* = Hedges' *g* (difference between groups in standard deviation units). Significant Mann-Whitney *U* tests are indicated with bold text. * = $p < .05$, ** = $p < .01$, ⁺ = $p < .001$, ° = $p < .1$.

§Unsat (Contact + Instrument Events), Archival vs Assigned VR group: Although the mean of the Assigned VR group is slightly higher than the mean of the Archival group, the

Event Modifier	Comparison	*M*(*SD*)	*n*	*U*	η^2	*g*

mean rank is lower (Archival Mean Rank = 458.40, Assigned VR Mean Rank = 347.27), indicating that the distribution of unsats is actually greater for the Archival group than for the Assigned VR group.

9.3. Appendix 3: Odds of Receiving an Unsat

Table 5. Fisher's Exact Tests on number of SNAs receiving at least one Unsat

Phase	Comparison	Number with "unsats"	Number without "unsats"	*p*	Odds Ratio
All Contact + Instruments Flights	**Omnibus**	-	-	**<.001**	-
	Archival vs Free VR	**223 88**	**613 2926**	**<.001**	**12.10**
	Archival vs Assigned VR	**223 1**	**613 63**	**<.001**	**22.92**
	Archival vs VIPER®	**223 0**	**613 52**	**<.001**	**38.25**
	Free VR vs Assigned VR	88 1	2926 63	1.000	1.89
	Free VR vs VIPER®	88 0	2926 52	.403	3.18
	Assigned VR vs VIPER®	1 0	63 52	1.000	2.48
All Contact Flights	**Omnibus**	-	-	**<.001**	-
	Archival vs Free VR	**211 83**	**625 2931**	**<.001**	**11.92**
	Archival vs Assigned VR	**211 1**	**625 63**	**<.001**	**21.27**
	Archival vs VIPER®	**211 0**	**625 52**	**<.001**	**35.50**
	Free VR vs Assigned VR	83 1	2931 63	1.000	1.78
	Free VR vs VIPER®	83 0	2931 52	.402	2.99
	Assigned VR vs VIPER®	1 0	63 52	1.000	2.48
All Instruments Flights	**Omnibus**	-	-	**<.001**	-
	Archival vs Free VR	**23 5**	**758 2501**	**<.001**	**15.18**

	Archival vs Assigned VR	23 0	758 62	.404	3.87
	Archival vs VIPER®	23 0	758 36	.619	2.26
	Free VR vs Assigned VR	5 0	2501 62	1.000	0.27
	Free VR vs VIPER®	5 0	2501 36	1.000	0.16
	Assigned VR vs VIPER®*	0 0	62 36	-	0.58

Note. Number with unsats = the number of SNAs who received at least one Unsat, Number without unsats = the number of SNAs who did not receive any unsats, *p* = 2-sided significance value for the Fisher's Exact Test, Odds Ratio = ratio of odds of receiving at least one Unsat. Odds ratio can be interpreted as "how many times more likely one group was to receive an Unsat."
*Assigned VR vs VIPER®, Instruments Phase: Fisher's Exact Test was not conducted, because no SNA had an Unsat in the Instruments Phase for these two groups.

9.4. Appendix 4: T-6B Curriculum Breakdown Survey

		Ground Training			
		Administration / Indoctrination	Systems	Operating Procedures	Course Rules
Current Media		CLASS	CLASS	CLASS	CLASS
VR/AR Devices	**Can VIPER support this training block?**	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe
	Notes				

		Contact Training					
		Contact Flight Procedures 1	Contact Flight Procedures 2	Contact Flight	Contact Cockpit Procedures	Contact Emergency Procedures Trainer	Contact
Current Media		MIL/CAI	MIL/CAI	LECT	UTD	UDT/OFT	OFT
VR/AR Devices	**Can VIPER support this training block?**	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe
	Notes						

		Contact Training				
		Day Contact	Midphase Contact Check Flight	Contact Solo Flight	Final Contact Check Flight	Night Contact
Current Media		T-6B	T-6B	T-6B	T-6B	T-6B
VR/AR Devices	**Can VIPER support this training block?**	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe
	Notes					
		Instrument Training				

		Instruments	Basic Instruments	Radio Instruments	Instrument Navigation	Instrument Check Flight
Current Media		CLASS	UTD	OFT & T-6B	OFT & T-6B	T-6B
VR/AR Devices	**Can VIPER support this training block?**	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe
	Notes					

		Navigation Training		
		Navigation (VFR)	Day Navigation	Night Navigation
Current Media		MIL/CAI	OFT & T-6B	OFT & T-6B
VR/AR Devices	**Can VIPER support this training block?**	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe
	Notes			

		Formation Training				
		Formation	Formation	Formation	Formation Solo Flight	Cruise Formation
Current Media		MIL/CAI	OFT	T-6B	T-6B	T-6B
VR/AR Devices	**Can VIPER support this training block?**	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe	Yes/No/Maybe
	Notes					

9.5. Appendix 5: T-6B VIPER® Maneuver Feedback Questionnaire

Instructions: Please respond to the following questions after observing or flying in the T-6B PTN VR-PTT with VIPER, the virtual instructor.

CONTACT

1. AILERON ROLL

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

2. APPROACH TURN STALL (ATS)

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

3. BARREL ROLL

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________

____

On this maneuver, is there any additional feedback you would like to provide? ______________________

____

4. CUBAN 8

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________

____

On this maneuver, is there any additional feedback you would like to provide? ______________________

____

5. GX

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? ______________________
____

6. IMMELMANN

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? ______________________
____

7. INTENTIONAL SPIN

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? ______________________
____

8. LANDING ATTITUDE STALL

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________

____

On this maneuver, is there any additional feedback you would like to provide? _____________________

____

9. LOOP

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1,2, or 3): ________________________________

____

On this maneuver, is there any additional feedback you would like to provide? _____________________

____

10. POWER OFF (ELP) STALL

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

11. POWER ON STALLS

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

12. SLOW FLIGHT

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

13. SPLIT-S

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

14. TURN PATTERN (TP)

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide?_____________________
____

15. UNUSUAL ATTITUDES RECOVERY (VMC) - CONTACT

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

16. WINGOVER

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1,2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

TAKEOFF/LANDINGS

17. ABORTED TAKEOFF

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide?______________________
____

18. ILS APPROACH

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? ______________________
____

19. LANDING PATTERN

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? ______________________
____

20. LOCALIZER APPROACH

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? ______________________
____

21. PRECAUTIONARY EMERGENCY LANDING (PEL)

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? ______________________
____

22. PRECAUTIONARY EMERGENCY LANDING IN PATTERN (PELP)

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? ______________________
____

23. STRAIGHT-IN APPROACH

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________

____

On this maneuver, is there any additional feedback you would like to provide? _____________________

____

24. TAKEOFF

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________

____

On this maneuver, is there any additional feedback you would like to provide?_____________________

____

25. WAVEOFF

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

INSTRUMENT

26. ARC AND RADIAL INTERCEPTS

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

27. ARCING

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

28. CONSTANT AIRSPEED CLIMBS

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________

____

On this maneuver, is there any additional feedback you would like to provide?______________________

____

29. CONSTANT AIRSPEED DESCENTS

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________

____

On this maneuver, is there any additional feedback you would like to provide?______________________

____

30. LEVEL SPEED CHANGES

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

31. RADIAL INTERCEPTS

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

32. STEEP TURNS

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________
____

On this maneuver, is there any additional feedback you would like to provide? _____________________
____

33. UNUSUAL ATTITUDES RECOVERIES (IMC) INSTRUMENT

	1 – Not Accurate At All	2 – Mostly Inaccurate	3 – Somewhat Inaccurate	4 – Somewhat Accurate	5 – Mostly Accurate	6 – Extremely Accurate
Monitoring of the aircraft						
Instruction prior to maneuver						
Feedback upon completion of maneuver						
	1 – Not Effective At All	2 – Mostly Ineffective	3 – Somewhat Ineffective	4 – Somewhat Effective	5 – Mostly Effective	4 – Extremely Effective
Effectiveness for student instruction						

Please explain any ratings of inaccurate or ineffective (1, 2, or 3): ________________________________

____

On this maneuver, is there any additional feedback you would like to provide? _____________________

____

34. General VIPER Utility and Usability. For usability, please consider the VIPER program only, not the VR device it is on.

	1 – Not At All Useful	2 – Mostly Not Useful	3 – Somewhat Not Useful	4 – Somewhat Useful	5 – Mostly Useful	6 – Extremely Useful
When asked a question, VIPER provides a useful response						
	1 – Not At All Helpful	2 – Mostly Not Helpful	3 – Somewhat Not Helpful	4 – Somewhat Helpful	5 – Mostly Helpful	6 – Extremely Helpful
For students, VIPER is more helpful than VR practice alone (w/o VIPER)						
	1 – Not At All Easy	2 – Mostly Not Easy	3 – Somewhat Not Easy	4 – Somewhat Easy	5 – Mostly Easy	6 – Extremely Easy
VIPER was easy to set up						
VIPER was easy to use after set up						
	1 – Not At All Reliable	2 – Mostly Unreliable	3 – Somewhat Unreliable	4 – Somewhat Reliable	5 – Mostly Reliable	6 – Extremely Reliable
I could rely on the VIPER software not to crash during use						

Please explain any ratings of 1, 2, or 3: ____

____

____

____

____

____

What are the most important changes/upgrades that could make VIPER more useful? ________________
____
____
____
____
____
____
____

Is there any additional feedback you would like to provide? ____________________________________
____
____
____
____
____
____
____

9.6. Appendix 6: T-6B VIPER® Questionnaire

Instructions: Please respond to the following questions after observing or flying in the T-6B PTN VR-PTT with VIPER, the virtual instructor.

NOTE: Your DODID is only being collected to track your survey data and device usage with performance in the aircraft. Your DODID will not be included with any raw data or data analysis sent outside of the research team.

Demographics

1. **Please circle your gender**: Male Female

2. **Please provide information about your current position:**
   - ☐ Student
     - Chapter of the syllabus: ____
   - ☐ Recent Graduate
   - ☐ Instructor
     - Instructor experience (Platform and Years): _______________________________
     - Flight experience (Platform and Years): __________________________________
   - ☐ Other: ____

3. **How many hours have you spent using VR systems in the past?:** _________________
4. **How many hours have you spend using virtual instructors in the past?**:____________

T-6B VR Introduction Overview

You were given an introduction to the VR systems and VIPER by an instructor prior to your practice sessions. The following questions are about the quality of the instructor-led overview.

5. **Was the introduction overview session with the IPs helpful in preparing you to interact with the VR devices?**

   a) Overall

1	2	3	4
Not Helpful at all	*Slightly Helpful*	*Very Helpful*	*Extremely Helpful*

   b) Purpose of device

1	2	3	4
Not Helpful at all	*Slightly Helpful*	*Very Helpful*	*Extremely Helpful*

   c) VR flight maneuvers

1	2	3	4
Not Helpful at all	*Slightly Helpful*	*Very Helpful*	*Extremely Helpful*

d) User guide reference

1	2	3	4
Not Helpful at all	*Slightly Helpful*	*Very Helpful*	*Extremely Helpful*

e) Technical support contact information (e.g., Engineer)

1	2	3	4
Not Helpful at all	*Slightly Helpful*	*Very Helpful*	*Extremely Helpful*

f) Scenario practice with instructor

1	2	3	4
Not Helpful at all	*Slightly Helpful*	*Very Helpful*	*Extremely Helpful*

6. What would you change to make the introduction overview more helpful?

____
____
____
____

T-6B VR Introduction Overview

You were given an introduction to the VR systems and VIPER by an instructor prior to your practice sessions. The following questions are about the quality of the instructor-led overview.

T-6B VR Training Curriculum

This section asks about how **VIPER** can support your current training curriculum. **Please only consider VIPER and not the VR device that VIPER is hosted on.**

7. Please select all the reasons that a student could use this device with VIPER (if any):

- ☐ Preparing for their next event
- ☐ Remediation on items for which their instructors gave feedback
- ☐ Learning new content
- ☐ Building a sight picture
- ☐ Other: ____

8. For each chapter, how effective is practice on this device with VIPER?

Contacts

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

If you selected *Not effective at all,* please provide an explanation:

____
____
____

Instruments

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

If you selected *Not effective at all,* please provide an explanation:

____
____
____

Navigation

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

If you selected *Not effective at all,* please provide an explanation:

____
____
____

Formation

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

If you selected *Not effective at all,* please provide an explanation:

____
____
____

9. Which chapters (if any) SHOULD NOT be practiced on this device with VIPER (i.e., using VIPER could hurt training outcomes)?

- ☐ Contacts:____
- ☐ Instruments: ____
- ☐ Navigation: ____
- ☐ Formation: ____

10. Do you have any additional feedback about how using these devices with VIPER can support the curriculum?

____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____

VIPER Utility

This section will ask you about how useful **VIPER** is as a training aid. **Please only consider the utility of VIPER, and not the utility of the VR device that VIPER is hosted on.**

11. How effective would VIPER be for students practicing the following maneuvers?

a) Takeoff

☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

b) Power on Stalls

☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

c) **Approach Turn Stall**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

d) **Landing Attitude Stall**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

e) **GX**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

f) **Steep Turns**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

g) **Level Speed Changes**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

h) **Landing Pattern**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

i) **ILS Approach**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

j) **Localizer**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

k) **Unusual Attitude Recovery (VMC)**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

l) **Unusual Attitude Recovery (IMC)**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

m) **Slow Flight**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

n) **Radial Intercepts**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

o) **Arcing**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

p) **Arc and Radial Intercepts**
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

q) Constant Airspeed Climbs

☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

r) Constant Airspeed Descents

☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

s) Waveoff

☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

t) Precautionary Emergency Landing (PEL)

☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

u) Precautionary Emergency Landing in Pattern (PELP)

☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

v) Turn Pattern

☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

w) Power Off Stall

☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

x) Aborted Takeoff
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

y) Intentional Spin
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

z) VFR Straight-In Approach
☐ N/A- did not use this maneuver

1	2	3	4
Not effective at all	*Slightly effective*	*Very effective*	*Extremely effective*

12. Is VIPER:

a) Giving you *timely* instruction for actions that you will perform next?
☐ Yes
☐ No
If not, when is it not? ____
____
____

b) Giving you *timely* feedback about actions you have completed?
☐ Yes
☐ No
If not, when is it not? ____
____
____

c) Accurately monitoring the state of the aircraft (e.g., its location, positioning)?
☐ Yes
☐ No
If not, when is it not? ____
____
____

d) Giving you *accurate* instruction for actions that you will perform next?
☐ Yes
☐ No
If not, when is it not? ____
____
____

e) Giving you *accurate* feedback about actions you have completed?
- ☐ Yes
- ☐ No

If not, when is it not? ____

____

____

f) As informative as it should be about upcoming actions?
- ☐ Yes
- ☐ No

If not, when is it not? ____

____

____

g) As informative as it should be about actions you have completed?
- ☐ Yes
- ☐ No

If not, when is it not? ____

____

____

h) Giving you accurate feedback in the after action review (if applicable)?
- ☐ Yes
- ☐ No

If not, what is inaccurate? ____

____

____

i) As informative as it should be in the after action review (if applicable)?
- ☐ Yes
- ☐ No

If not, what would make it more informative? ____

____

____

13. When you ask VIPER a question, does it provide a useful response?
- ☐ I have never asked VIPER a question
- ☐ Yes

  What makes it useful?____

  ____

  ____
- ☐ No

  What would make it more useful?____

  ____

  ____

14. VIPER's feedback about actions you have completed is consistent with feedback given by human instructors.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

15. VIPER's advice about upcoming actions is consistent with advice given by human instructors.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

16. I can make accurate decisions based on the guidance that VIPER provides about upcoming actions.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

17. I can take accurate actions to correct my mistakes based on the feedback that VIPER provides.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

18. How effective is VIPER compared to VR alone?

1	2	3	4	5
Much less effective	*A little less effective*	*Equally effective*	*A little more effective*	*Much more effective*

19. How effective is VIPER compared to in-person instruction?

1	2	3	4	5
Much less effective	*A little less effective*	*Equally effective*	*A little more effective*	*Much more effective*

20. I am confident that VIPER will improve students' performance in the aircraft more than VR without VIPER.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

☐ I don't know

21. VIPER could increase mission readiness more than the VR without VIPER.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

☐ I don't know

22. How motivated are you to use VIPER?

1	2	3	4
Very unmotivated	*Slightly unmotivated*	*Slightly motivated*	*Very motivated*

23. Would you recommend VIPER to future students?

☐ Yes
☐ No

Why?____
____
____

24. **Do you have any additional feedback about VIPER's utility?**

____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____
____

VIPER Usability

This section asks about **VIPER's** reliability, functionality, and ease of use. **Please only consider the usability of VIPER and not the usability of the VR device that VIPER is hosted on.**

25. **Is VIPER's "how to" *video* effective in teaching you how to use VIPER?**
    - ☐ I did not view the "how to" video
    - ☐ Yes
    - ☐ No

    If not, why is it ineffective? ____
    ____
    ____

26. **Is VIPER's "how to" *training maneuver* effective in teaching you how to use VIPER?**
    - ☐ I did not use the "how to" training maneuver
    - ☐ Yes
    - ☐ No

    If not, why is it ineffective? ____
    ____
    ____

27. **Does using VIPER cause any delays in the aircraft's response to your inputs (e.g., selecting flap position is quickly reflected in aircraft performance)?**
    - ☐ Yes
    - ☐ No

    If so, when? ____
    ____
    ____

28. Does using VIPER cause any errors in the aircraft's response to your inputs?

- ☐ Yes
- ☐ No

If so, when? ____

____

____

29. VIPER was easy to set up.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

30. VIPER was easy to use after being set up.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

31. VIPER was distracting during my practice sessions.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

If so, what part was distracting? ____

____

____

32. VIPER's text was clear enough to read.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

33. VIPER's text was in an appropriate location for easy access.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

34. VIPER's auditory instructions/feedback were clear enough to understand.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

35. VIPER provided information in a way that I could understand.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

36. I could rely on VIPER to start up without crashing.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

37. I could rely on VIPER to run through my entire practice session without crashing.

1	2	3	4
Strongly disagree	*Disagree*	*Agree*	*Strongly Agree*

38. Do you have any other feedback about the usability of VIPER?

____

9.7. Appendix 7: VIPER® VR-PTT Practice Log Book

Please write and bubble in your 10-digit DODID.

⓪	⓪	⓪	⓪	⓪	⓪	⓪	⓪	⓪	⓪
①	①	①	①	①	①	①	①	①	①
②	②	②	②	②	②	②	②	②	②
③	③	③	③	③	③	③	③	③	③
④	④	④	④	④	④	④	④	④	④
⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤
⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥
⑦	⑦	⑦	⑦	⑦	⑦	⑦	⑦	⑦	⑦
⑧	⑧	⑧	⑧	⑧	⑧	⑧	⑧	⑧	⑧
⑨	⑨	⑨	⑨	⑨	⑨	⑨	⑨	⑨	⑨

DODID Verified by: ______________________________

Date	Scenario Name	Event(s) Preparing for	Start Time	End Time	Reps	Focus of Training
						○ Self Study ○ Event Preparation ○ Event Remediation
						○ Self Study ○ Event Preparation ○ Event Remediation
						○ Self Study ○ Event Preparation ○ Event Remediation
						○ Self Study ○ Event Preparation ○ Event Remediation
						○ Self Study ○ Event Preparation ○ Event Remediation
						○ Self Study ○ Event Preparation ○ Event Remediation
						○ Self Study ○ Event Preparation ○ Event Remediation

9.8. Appendix 8: VIPER® Wrap Up Questionnaire

Overall Usability:

1. Have you explored the capabilities of the VIPER program?
    a. What do you think the benefits are?
    b. What do you think the limitations are?
2. Are there basic **operability** issues with VIPER? This refers to getting the program started, using the program for practice, calibration, programming, etc.
    a. Major **operability** issues (i.e., must be fixed)?
    b. Minor **operability** issues (i.e., "nice to haves")?
3. Did the dashboard make it easy to create/edit event profiles on your own?
    a. If not, what was challenging?
4. Was the voice recognition software useful for practice (i.e., being able to request a maneuver to practice, demo, etc.)?
    a. What were the benefits to utilizing voice recognition?
    b. What were the limitations to utilizing voice recognition?
5. Were the maneuvers available for practice within VIPER useful to students? Why or why not?
6. To your knowledge, was VIPER ever recommended to any students outside of the scope of the research study? For example, to assist with practice on specific maneuvers.
    a. If yes, what was VIPER recommended for?
    b. If no, why do you think VIPER was not recommended?

VIPER Coaching & Feedback:

7. Is the feedback VIPER provides during a practice session consistent with what a live instructor would provide (i.e., verbal questions and coaching upon completion of a maneuver)?
    a. What were the benefits of VIPER's verbal feedback?
    b. What were the limitations to VIPER's verbal feedback?
8. Did the dashboard displaying an overview of an individual's practice provide useful information on performance and progress?
    a. What were the benefits of the dashboard?
    b. What were the limitations of the dashboard?
9. For students utilizing VIPER, did you notice any changes in student performance or knowledge?
    a. What were the changes?
    b. Were the changes good and/or bad? Why?

VIPER Implementation:

10. Do you have any recommendations regarding best practices for use of VIPER as part of the training curriculum? (e.g., study-only, scheduled events, specific phase or stage of training, etc.)
11. Overall, do you think instructors and students would be receptive to the use and integration of VIPER in training? Why or why not?
12. Are there modifications or improvements that could be made to influence the willingness to adopt and integrate VIPER as part of training?

9.9. Appendix 9: List of Abbreviations and Acronyms

BMT	Behavioral Modeling Training
CNATRA	Chief of Naval Air Training
COTS	Commercial Off-The-Shelf
DMI	Discovery Machine, Inc.
IP	Instructor Pilot
ITD	Immersive Training Device
MXR	Multidisciplinary Extended Reality
NAS	Naval Air Station
NATN	Naval Aviation Training Next
NAWCTSD	Naval Air Warfare Center Training Systems Division
NIFE	Naval Introductory Flight Evaluation
OFT	Operational Flight Trainer
PMA-205	Program Management Activity-205; Naval Aviation Training Systems and Ranges Program Office
PTN	Pilot Training Next
SAIC	Science Applications International Corporation
SNA	Student Naval Aviator
T-SHARP	Training Sierra Hotel Aviation Readiness Program
UTD	Unit Training Device
VIPER®	Virtual Instructor Pilot Exercise Referee®
VR	Virtual Reality
VR-PTT	Virtual Reality Part-Task Trainer