FITS Combined Task 1&2

Final Report

For the Embry Riddle Aeronautical University Effort

Prepared for:
M. Summers
On Site FITS Director
Embry Riddle Aeronautical University

Prepared by:
Jon French
Beth Blickensderfer
Frank Ayers
Tom Connolly

College of Arts and Sciences
College of Aviation
Embry Riddle Aeronautical University
Daytona Beach, FL 32114

August, 2005

FITS Combined Task 1&2

Final Report

Table of Contents

SECTION		PAGE
Executive Summary	……………………………………	3
Background	……………………………………	4
Method	……………………………………	11
Results	……………………………………	21
Discussion	……………………………………	30
References	……………………………………	33
Appendices	……………………………………	A-1

FITS Combined Task 1&2
Final Report

EXECUTIVE SUMMARY

This report describes the results of the combined Tasks 1 and 2 for the FAA Industry Training Standards (FITS) program conducted at Embry Riddle Aeronautical University. These tasks required a comparison of traditional instrument training, called maneuvers based training (MBT) in this report, and scenario-based training (SBT) which forms the basis of the FITS recommended procedures for technically advanced aircraft (TAA). These two instrument flight instruction methods were assessed using subjective questionnaires regarding workload, situational awareness, self-efficacy, and decision-making skills, as well as expert evaluation of instrument piloting and navigation performance in a controlled double blind statistical paradigm.

All 27 participants were instrument rated pilots with less than 500 total flight hours and virtually no experience with TAA. All were randomly assigned to one of the two conditions, MBT or SBT. All received 8 hours of training before the final posttest evaluation. The primary TAA simulation device was a Cirrus SR20 with a Chelton primary flight display (PFD) and multi-function display (MFD) powered by Micro Soft Flight Simulator. The participants did not know they were taking part in an evaluation of FITS.

An experimentally blind rater scored instrument flight skills during a pre-training (pre-test) flight and a similar post training (posttest) data collection flight. These were rated for 8 phases of flight (such as Take off, en route navigation) and subjective questionnaires were completed after both flights. On most of the measures, a significant improvement was found between pre-test and posttest measures, indicating the effectiveness of both training methodologies. In addition, the SBT group performed statistically better on 5 of the 8 measures of piloting ability than the MBT group for the posttest flight. In no case did the SBT group score worse than MBT. Further, the SBT group demonstrated a tendency to report reduced workload and an improvement in self-efficacy and situational awareness compared to the MBT.

The results indicate that scenario-based training may lead to improved piloting and navigation skills over traditional maneuvers based training techniques for TAA flight. Although MBT is a reliable method to teach flight and has been used successfully for decades, FITS may be the perfect opportunity to exploit advances in teaching methodologies to accompany the advances in automated GPS based instrument flight. There are several additional questions raised by these results that should be addressed by future research; for example, over how many months will these improvements be maintained? Also, what other measures of pilot efficacy would be improved by the training, such as scanning techniques, for the TAA? In a carefully controlled laboratory setting, SBT was shown to be as good as, often better and never worse, than MBT in teaching TAA flight.

FITS Combined Task 1&2: Final Report

OBJECTIVE

This is the Final Report for the combined Tasks 1&2 of the FAA Industry Training Standards (FITS) program. It describes an evaluation of the FITS methodology for supplemental flight instruction using novel 'glass' instruments in technically advanced aircraft (TAA). This report briefly explains the FITS approach, thoroughly summarizes the methods, the results obtained, and the implications of the Task 1&2 research conducted at Embry Riddle Aeronautical University.

Task 1&2 were combined in Fall 2004 because both tasks focused on quantifying the effectiveness of scenario-based training compared to traditional instrument training for TAA. Task 1 focused on preparing the simulated flight performance research with literature reviews and research plans, and Task 2 concentrated on completing the research project. Since these two tasks concentrated on the same research problem, they were combined.

Embry Riddle Aeronautical University (ERAU) and the University of North Dakota (UND) partnered in planning for the research that would evaluate the effectiveness of the FITS approach. A Literature Review Report and a Metrics Plan (the outcome measures to be used to determine effectiveness) were prepared and submitted as Deliverables. Both are included as the first and second items in accompanying Appendices.

The original 2004 plan called for data collection at ERAU to exactly parallel those collected at UND. In that way, the data could be combined to reduce the number of participants (subjects) at both universities. However, largely due to differences in the flight simulation device available at the two universities, it became apparent that the majority of the data could not be combined due to considerable procedural differences. The major disadvantage of this outcome was a reduction in the statistical power that might be needed to demonstrate the effectiveness of FITS training. It was hoped that the FITS training might show a large enough improvement over traditional methods that a robust effect of the training could be expected in at least some of the outcome measures. These measures then, would not suffer from a reduced sample size. This turned out to be the case. Accordingly, this report will consist of only data collected from the ERAU effort. The UND data will be provided under a separate cover.

The goals of the combined FITS Task 1&2 were to evaluate the effectiveness of the FITS training in the following areas:

- Improvement in judgment and decision making skills
- Improvement in automation management skills
- Improvement in situational awareness
- Identify differences in FITS and non FITS trained pilot skills
- Identify specific weaknesses and strengths of the FITS program.

BACKGROUND

Due to an increased demand, particularly in small, general aviation type aircraft, there is a growing proliferation of glass-instrumented aircraft or TAA. These digital microprocessor controlled instruments represent a dramatic advance in avionics over the traditional, pressure driven analog instruments that have been in existence since the advent of the airplane. Ordinarily the TAA consists of two displays; a primary flight display (PFD), that describes the flight characteristics of the aircraft (heading, altitude, attitude, etc), and the multifunctional flight display (MFD) that is a moving map navigational aid. Thus, for purposes of this study, a TAA may be considered to be any aircraft with advanced (computerized) flight management navigation system that couples a global positioning satellite (GPS) link with an autopilot. These advanced instruments cost less than traditional ones, are easier to interpret, and are more reliable. They represent an order of magnitude improvement in avionics and quickly found their way into commercial aviation. They are viewed as having the ability to reduce workload, improve piloting and navigation, and even to improve pilot judgment and decision-making ability. They also have advanced computerized flight and navigation features to understand so supplemental training is needed for general aviation (GA) pilots: this is the focus of the FITS program. Since GA TAA might compromise flight safety in the national air space, particularly because training demands are far from what is required in commercial aviation, the FAA acknowledged that additional training was needed for GA TAA pilots.

The FITS program is an effort to improve the training that pilots receive for TAA flight. Traditional instrument training relies on learning basic maneuvering skills and has served aviation well for decades. It was developed over the early years of aviation and sometimes fine-tuned by the high cost in life and material born of trial and error. Over the years, there have been considerable advances in education and, with the advent of TAA, a leap forward in digital avionics. This report is based on the premise that it is the right time to take advantage of the advance in avionics with an accompanying advance in TAA training. The FITS partners quickly settled on a new approach to TAA instrument training, one that has worked successfully in other high technology domains, like medicine and military sensor training. This approach utilizes scenario-based training in which the pilot is required to design and practice dozens of flight scenarios, from take-off and enroute navigation to flight emergencies and landing. The FITS program also emphasizes the involvement of the student in his knowledge and skills acquisition. The student is required to identify his own deficiencies for each scenario and create a flight scenario to advance needed skills. In addition, FITS syllabi have been developed by drawing from military, academic and industry programs that identify skill sets and training standards that should be included for most types of pilot training, from ab initio to recurrent very light jet aircraft and more. Thus FITS is both comprehensive pilot training programs and student self-evaluations.

Traditional instrument flight uses a skill based approach and can be considered a task based approach. Pilots are trained on particular maneuvers to acquire skills in these to the point that they produce an almost automatic response. It will be called maneuvers based training (MBT) for the purposes of this report. MBT emphasizes the skills required of the FAA's Practical Test Standards (PTS) and over many years may have drifted towards a practice of teaching to the test. In addition, the skills are sometimes learned in isolation; for example, one learns the standard soft field landing technique rather than coupling that training with a condition or scenario in which it might arise during flight. MBT training is by example, the instructor's example, and the student mimics the maneuver until accomplishing it successfully, as determined by the instructor. It is rarely tied to any use other than the practice, and the student is passive in the process, not developing the ability to identify and correct weaknesses. The FITS partners designed a program to address these deficiencies.

In scenario-based training (SBT), the essence of the FITS approach, the student takes the lead in designing scenarios for practice and in evaluating his performance. SBT can also be called event-based training. The idea is to generate a self-evaluation process that continues beyond the class as the pilot self-trains throughout his aviation career. The student learns the same skills and maneuvers as in MBT but these are practiced in the context of a scenario, with many scenarios coupled to the maneuver until the student has not only the requisite skills but can relate them to many conditions where they would be needed. The goal of SBT is that when a condition occurs requiring the maneuver, the SBT pilot is more likely to respond correctly, since something similar to the condition has been faced before and faster than the MBT pilot, who must search his memory to link a maneuver to a situation.

The idea underlying SBT is to give the learner opportunities to acquire knowledge and skills necessary for task performance via simulated "real-world" operational environments. The required competencies are trained within the scenarios, as well as upon completion via instructor feedback (Oser, 1999). Thus, active learning, extensive practice and feedback are the cornerstones of SBT, and these are also the characteristics that distinguish SBT from other training methods (Salas & Cannon-Bowers, 2000).

The procedures of SBT are not new to training operators of complex and sophisticated equipment; they are just new to flight training because an organized, large-scale effort has never been made to formalize the procedure. For quite some time training researchers and practitioners have known the value of using simulated tasks for skill development (Goldstein, 1993). The idea underlying SBT is to give the learner the opportunity to acquire the knowledge and skills necessary for task performance via a simulated "real-world" operational environment. The required competencies (e.g., operator knowledge, skills, and attitudes) are trained within the scenarios, as well as upon completion via instructor feedback (Oser, Cannon-Bowers, Salas, & Dwyer, 1999a). Thus, active learning, extensive practice and feedback are the cornerstones of SBT (Oser et al., 1999a).

The SBT Cycle

SBT is more than learners simply practicing simulated tasks: little, if any use of lectures, books, and class discussions occurs. Thus, in SBT, the *scenario is the curriculum*. Therefore, it is essential that SBT offer a highly structured training which links learning objectives, exercise design, performance measurement and feedback (Dwyer, Oser, Salas & Fowlkes, 1999). As shown in Figure 1, the process begins with the instructor establishing the desired *skills to train* based on the *training objectives* (Oser et al., 1999a). Training objectives generally arise from a need for improvement or because unfamiliar equipment is introduced. Based on the required skills, the instructor plans s*cenario events* that will give the learner the opportunity to practice the desired skills. If planned carefully, the events ensure the learner practices all of the targeted skills until all training objectives are met (Cannon-Bowers, Burns, Salas & Pruitt, 1998).

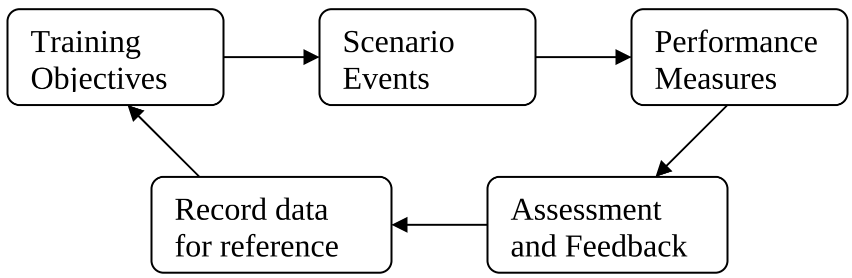

Figure 1. Scenario-based training model, adapted from Oser et al. 1999a

The next step in the process is *performance measures* (Oser et al., 1999a). Ideally, criteria closely linked to scenario events are established for instructors to assess whether the learner has developed the necessary knowledge and skills. Once the learning objectives, scenario events, and performance criteria are established, it is time for the learner to experience the scenario. During this time, the instructor monitors the learner's performance and uses the performance criteria to judge the learner's performance. The next step in the SBT cycle is *feedback/debrief*. In this step, the instructor gives specific, detailed feedback to the learner. Ideally, the instructor links the feedback to specific events and learning objectives. This helps the learner to interpret the feedback and, in turn, to integrate the feedback into his/her subsequent performance. The final step, *record data for reference*, simply means that the results of the training can and should be used when identifying future training objectives.

In sum, the SBT method offers a structured approach, using practice and feedback, to train learners the way they will perform on the job.

Scenario-based training has been applied in a variety of domains. It is difficult to discern, however, the degree to which the tightly linked SBT cycle described above was followed. Furthermore, only limited data appears. Nevertheless, a quick review of some of the domains in which a version of SBT has appeared is warranted.

To begin, results indicate that SBT has high face validity—the learners believe it to be an effective training method. For example, stressful and realistic scenarios were presented in a training program for firearm safety with the goal to increase mental readiness of probation officers (Scharr, 2001). In a post-training evaluation questionnaire, 97% of the respondents felt the training was effective "to a great extent." Additionally, Lowry

(2000) found that probation officers trained with SBT were more likely to rate the training as excellent than were officers trained with other methods.

Training cost savings also have been reported. Using a method similar to SBT, Stewart, Dohme & Nullmeyer (2002) examined the effectiveness of pilot training for rotary wing aircraft. Training times and costs were reduced as a result, and the transfer of training ratio was found to be acceptable. These results are not surprising, as Cannon-Bowers et al. (1998) described: training time and cost are reduced as a result of including only those events that exercise the targeted skills.

In addition, a network of simulators was used to train military service personnel (i.e., U.S. National Guard, U.S. Marine Corps, U.S. Army and U.S. Air Force) around the world in a joint military exercise. In this exercise, Dwyer et al. (1999) tested the use of performance measurement tools designed to specifically link to events and, thus, learning objectives. Although only case studies were used, the results indicated the SBT approach was successful.

Finally, while no data were reported, scenario-based training has been used in the medical field to allow doctors to practice techniques such as sigmoidoscopy (Kneebone et al., 2003). Using a simulated patient, participants were able to improve their procedural skills without risking harm to a live patient.

Thus, SBT has been used in a variety of domains with successful results. Additional research on SBT effectiveness, however, would enable training designers to take maximal advantage of the method while avoiding pitfalls. The next section of this paper describes the introduction of the technologically advanced aircraft (TAA) in general aviation and the potential use of SBT to train pilots to use the new system.

Since the 1970's, aircraft manufacturers have attempted to counteract the negative effects of increasingly complex aircraft and traffic-congested skies (NASA Facts Online, 2000). This effort began with the use of cathode ray tube screens in some aircraft to reduce the number of instruments and displays within the cockpit (NASA Facts Online, 2000). Electronic flight instrument-display systems (EFIS) (i.e., the "glass cockpit"; technically advanced aircraft) appeared a short time later (Mulder, van Paassen, & Mulder, 2004).

While EFIS systems differ depending on the manufacturer, the systems currently used in general aviation include complete flight/navigation instrumentation in glass cockpit technology. Featuring 3D synthetic vision and highway-in-the-sky on a LCD display, the flight display systems relay primary flight information as one integrated picture. The 3D synthetic vision enables pilots to see the terrain and future paths, while the highway-in-the-sky "creates a 3-D tunnel for all enroute and instrument procedures" and reduces pilot workload in executing flight plans and instrument approaches (Chelton Flight Systems, n.d.).

In commercial aviation, the new technology engendered initial concerns with automation -- e.g., how to keep the pilot active in the process and able to explain what the automation was doing (Mulder et al., 2004). However, over time, accident statistics have shown an increased level of safety. Presently, most in the aviation community would agree that the glass cockpit has “increased pilot safety by reducing pilot workload at peak times, yet kept the pilot ‘in the loop’ at all times to maintain situational awareness” (NASA Facts Online, 2000).

With the advent of the TAA in general aviation, industry, government, and academia are collaborating on the development of training guidelines to transition pilots to the new aircraft. These guidelines were developed by the FITS partners and have the overall goal to ensure that pilots learn how to fly a small jet aircraft “safely, competently and efficiently” (FAA/Industry Training Standards, 2004). Specific goals include training higher order thinking, procedural knowledge, and the related knowledge, skills, and attitudes necessary to reduce the number of aviation accidents attributed to pilot error.

A number of reasons exist as to why SBT was selected by FITS. Most importantly, scenario-based training is generally used for training complex tasks or interactions (Loftin, Wang, & Baffes, 1989). Complex tasks are those that require high-level cognitive processing skills such as aeronautical decision-making and situational awareness. Similar to the complex operational environments described by Oser et al. (1999b), aviators must continually identify cues and patterns in their environment, assess and make inferences about their status, make judgments about their next course of action, implement actions, and evaluate the impact of these actions. Training must therefore address successful performance of such complex processes and enhance the associated high order skills. As SBT replicates the critical aspects of the operational environment while also offering carefully planned events and feedback, it seems a logical choice to train such skills.

In contrast, traditional aviation training typically requires pilots to practice one flight task at a time until that task was mastered. The tasks are not integrated into a seamless activity until the training is near completion. This method may be appropriate for the mastery of some simple tasks, such as learning how to control the rudder, but is too narrow/limited for more complex tasks. The complex interactions between pilots, instrumentation, and ATC that arise when flying a TAA may require an alternative training method. Thus, focus has shifted to using SBT for training pilots to operate the TAA (FAA/Industry Training Standards, 2004).

Additionally, as the TAA is introduced to general aviation, instructors will be tasked with training aviators who have a wide range of prior experience. Again, SBT seems a good fit: a wide range of operators may benefit from SBT, from novices without any specialized knowledge to expert users who wish to familiarize themselves with a new product (Loftin et al., 1989).

With this background, the approach used for fulfilling the requirements of Task 1&2 will be presented.

The combined Task 1&2 presented two major problems that had to be accomplished before data could be collected. First, since there is no traditional instrument training for technically advanced aircraft (TAA), the team had to develop a maneuvers based training that would be applicable to TAA. Discussions with several TAA trained pilots and subject matter experts (SME) made it evident that many pilots are trained in TAA instrument flight using MBT principals, although not in a rigorous, consistent manner. This aided in the development of a well-defined program of training that relied on simple maneuvers rather than scenarios to accomplish glass instrument (TAA) flight training that was used in this experiment. A team of certified instrument instructor pilots who had no experience with SBT were selected to assist in developing this syllabus. It is included in the accompanying Appendix 23. Second, the team had to demonstrate the effectiveness of scenario-based training (SBT) compared to MBT. This was the real focus of Task 1 and 2 and was possible once the MBT glass instrument-training syllabus was developed for comparison. The same simulation device was then used for MBT and SBT of the participants to allow a direct comparison of training efficacy without confounding due to different simulation devices.

In discussions with experts at the Civil Aerospace Medical Institute (CAMI) in Oklahoma City, it became apparent that the majority of General Aviation (GA) pilots probably learn to fly TAA currently without formal training. Instead these GA pilots learned what they can from the aircraft system manuals. A few forward thinking aircraft manufacturers provided training in TAA flight but pilots are not required to learn or demonstrate proficiency in glass instrument operation before flying away with one. The TAA is designed to be far easier to use than analog instruments but a considerable amount of experience is essential and, as always, flight safety increases with the experience of the pilot in TAA. Therefore, it was decided to include this third group, essentially a No Training (NT) group with the SBT and MBT groups for comparison. However, a failure of the simulation device used in the study near the last month of the study prevented the completion of this group and the SBT group. The SBT group had 12 of the 15 pilots but the NT group only 9. Therefore, since many of the assumptions of a normal or Gaussian distribution would be violated in any statistical comparison, obfuscating the results for all the groups, the NT group was not reported on in this document. An amended document will be provided at a later date to complete this investigation. While unfortunate, the original Task 1&2 required the primary focus to be on the MBT and SBT comparisons.

This report details the results of a carefully controlled fight laboratory experiment that compared SBT to MBT on a number of skills and attitudes important to instrument flight in a TAA simulation device. It was expected that SBT pilots would measurably improve instrument-piloting measures over pilots who received MBT.

METHODS

The research was conducted in the Technically Advanced Aircraft Performance (TAAP) laboratory for pilots located in the Human Factors and Systems Department research facility at Embry Riddle Aeronautical University in Daytona Beach, FL. The experiment was conducted as collaborative effort between the aviation subject matter experts at the University's College of Aviation and the human performance assessment experts in the College of Arts and Sciences.

The ERAU team assembled to complete the Combined Task 1&2 consisted of the scientific/technical team, Drs. Jon French and Beth Blickensderfer and graduate students Angie Baskin and Shayna Strally; the subject matter experts, Dr Tom Connolly, Professors Frank Ayers and Michele Summers; the raters, graduate students Darren Dodd and Amanda Hendricks and the Instructor Pilots, Cynthia West, Michael Poulin and Nate Warrick. The contributions of each of these individuals will be obvious from the explanation of the methods below.

Originally, ERAU intended to have 15 participants in each of three experimental conditions; SBT, MBT and NT. This would have provided 45 participants; however, only 36 participants finished the training before a catastrophic failure of the Chelton simulation system used at ERAU forced early termination of the study. The cause of the equipment failure is a loose wire somewhere within the interface. Considerable effort has already been made to repair the equipment and the outcome will be a better system with upgrades to all the software involved and less prone to failure. Because only 9 of the NT group data were collected, it was decided that the inclusion of slightly more than half of the anticipated subjects would produce statistical anomalies that would be difficult to explain and would divert attention from the original intention of Task 1&2, that of comparing MBT to SBT in terms of pilot effectiveness. It is hoped that the remaining 9 participants can be completed soon to allow for the increase in statistical power that equal numbers in each condition brings before the combined Task 1&2 papers are published.

The breakdown of the training methods for the MBT and SBT conditions are explained below in the Procedures section. Neither the participants nor the raters were aware of which training methodology was used for each participant so the study utilized a double blind procedure.

PARTICIPANTS

Participants were kept unaware (experimentally blind) that the project was an FAA Industry Training (FITS) study as much as possible to prevent them from researching the specifics of the study from the publicly available FITS website, thereby learning the differences between SBT and MBT or worse, the conditions of the study. Instead, they were told they were participating in a Technologically Advanced Aircraft Performance (TAAP) study for pilots, and all questionnaires and forms reflected the TAAP headline. Every effort was made to keep them from talking to other pilots who might also participate in the study.

The 36 participants were randomly distributed to conditions as shown in Table 1. They were solicited from the general population of ERAU students and had to fit the general requirements of having less than 500 total flight hours, IFR currency and none or very little experience with glass instruments for inclusion. Two large calls for pilots resulted in over 100 volunteers and more than half of those met the inclusion criteria. The names of qualified participants that resulted were randomly assigned to one of the three conditions, NT, MBT or SBT. Pilots who were selected for inclusion were compensated $100 for their time upon their completion of the study.

Table 1. Characteristics of the Task 1&2 participants by condition. Averages are shown ± the standard error of the mean.

GROUP	CONDITION	Number of Participants	Average AGE	GENDER	Average Hours VFR	Average Hours IFR
1	NT	9	23 ± 2	9 Males	223 ± 30	42 ± 20
2	MBT	15	21 ± 1	14 Males	167 ± 16	42 ± 9
3	SBT	12	21 ± 1	11 Males	217 ± 40	53 ± 13

NT=No Training
MBT = Maneuvers Based Training
SBT = Scenario-based Training

General aviation and related information on each participant was collected on the Demographics Data Flight Performance I (DDFP-I) form, the Demographics Data Flight Performance II (DDFP-II) and the Technically Advanced Aircraft Instrument Training (TAA-IT) forms. Examples of these forms are provided in Appendices 5, 8 and 18, respectively. The Participant Data Form Checklist, also included (Appendix 3), lists the order in which these forms and all forms were given during the experiment. A printout from an Excel Spreadsheet is included in the Appendix 20 that details all the responses to the three demographic forms.

Demographic information from the participants further indicated that none of the 36 had any record of aviation related accidents, incidents or FAA violations. Further, as shown in the Demographics Spreadsheet in Appendix 20, the participants were fairly equal in their self-proclaimed ability to use computer hardware and software and most used a desktop computer daily. Very few indicated they used a computer for recreational pursuits, such as computer games, with any regularity. However, most estimate they had a good working knowledge of computer hardware and software.

Most of the participants indicated that they had received some formal autopilot training in their flight careers, all indicated that the use of the autopilot is a good idea and the great majority indicated they felt that the study seemed to emphasize autopilot use, independent of NT, MBT or SBT conditions.

The order the participants were trained was not random, although each of the originally qualified volunteers had an equal chance of being in any of the groups. The order of training was determined by the availability of instructors for the MBT and SBT groups.

Since the NT group was not considered in the analysis for this report, as explained above, only the MBT and the SBT groups will be described. For these two groups, all experimental conditions were exactly the same with the obvious exception of the training syllabus that was used.

The MBT consisted of the individual skill items listed in the Traditional Transition Syllabus for this training in Appendix 23. The SBT consisted of the instructor using the Scenario-based Training Syllabus that can be found in Appendix 24. Neither of these syllabi should be considered useful for general aviation. They were developed for research purposes only and are included to allow for a thorough review of the procedures used in the study.

INSTRUCTORS

The Instructors had the responsibility of training the participants in either the MBT or the SBT methods. They were required to be CFII and to be unfamiliar with (i.e. no formal training nor experience) TAA or glass instrument flight. ERAU has a large population of flight instructors from which to draw and the SMEs selected the most qualified to be MBT instructors. They were kept experimentally blind to the conditions of the study as much as possible for MBT training. They were trained in a classroom setting by the SMEs and the training was videotaped in the event other MBT trainers were needed.

The participants in group 1 (NT) did not require a trainer and so they were evaluated throughout the entire 9 months of data collection with about half collected before group 2, the MBT condition, had completed testing. The remainder of the NT group was completed as possible in parallel with the testing of the MBT and SBT trained pilots. This served to keep the raters (explained below) experimentally blind to which group they were rating, never knowing if the participant was an NT, MBT or SBT pilot. MBT group participants were tested before group 3 (SBT) primarily because of the quality of the instructor pilots. The three IPs who did all the training as MBT flight instructors also did the instruction for the SBT group. New instructors for the SBT group of the same caliber and background as our MBT group instructors would have been difficult to find and would have required training in the system and procedures of the study. It was decided to use the same instructors in the SBT group that we had used for the MBT group in order to maintain the same consistency of instructor quality and to reduce the opportunity for instructor bias to influence the results. The instructors quickly learned SBT training.

Once they became SBT instructors, there was no longer any need to keep them experimentally blind to the conditions since they did not interact with the raters who tested typically on Fridays (pre-test) and Mondays (posttest). The Instructors typically trained the individual participant for four hours on Saturday and four hours on Sunday giving 8 hours of formal instruction. This is after the participants had been given the pre-test by the raters, usually four hours on Friday and before the participants had been given the posttest as, usually four hours on Monday, as explained below.

MATERIALS

Ratings of performance of the pre and post test scenarios were used as the primary data for the experiment. The rating forms are shown in Appendix 12 and 16 for the pre and post test, respectively. These forms were developed by subject matter experts Frank Ayers and Tom Connolly, both very experienced pilots and educators. Their intent was to create key checklist items for every phase of flight particularly as they relate to glass instrumented flight. The pre-test was supposed to parallel the post test in item specificity, differing only in location or altitude or heading or some other non-training-essential feature. For the event labeled Take-off and Departure (Event 4) for example, Vero Beach was the departure point for the pre test and Daytona Beach for the post test but otherwise this event was similar in the post test for both in terms of weather, fuel and passenger, all the principle details, as for the pre test. The similarities of the event requirements were further enhanced by the raters during their training for inter rater reliability as explained below.

Both the pre-test and the post test considered the same piloting skills, as can be seen from Appendices 12 and 16. There were 8 phases of flight that were common to both:

Event 1: Flight planning
Event 2: Pre-Flight Preparation
Event 3: Pre-Take-off
Event 4: Take-off and Departure
Event 5: Re-Route
Event 6: En-Route
Event 7: Approach
Event 8: Missed Approach

Only the locations and specific events of the scenarios were different. Each of these events had numerous rating criteria. For example, for Event 3: Pre-Take-off, the pilot was rated as to whether he performed the engine run-up, set the avionics for flight, briefed the passenger and requested to be cleared to the active runway, among other criteria. For statistical evaluation purposes, the criteria for each event were compiled to give an event score rather than a criteria score. Further, a percent of those criteria that were correct for each event was used as the final score for each event. Thus, if a pilot got 3 out of 6 criteria correct for an event for example, that pilot’s score would be 50%. These were the data that were entered into statistical evaluation considered in the Results section below.

RATERS

Two raters were responsible for the pre and posttesting of all participants. They helped to develop the rating forms, along with the subject matter experts, that were used in the pre and posttests so they were very familiar with the intent and content of the rating forms before the experiment began. Rigorous efforts were made to keep the raters experimentally blind to the type of training each participant received. This was possible because the instructor training was typically done on the weekend, as explained above. They were both instrument rated and had received no formal training in SBT but had some familiarity with the concept from their classes. They both received thorough training in the use of the TAA simulation device and the procedures of the study prior to the beginning of data collection. To establish inter rater agreement, both raters evaluated the same 5 participants near the beginning of the experiment. The percent agreement was then calculated and, depending on the particular item being rated, agreement ranged between 65 and 100 percent. This check on inter rater agreement established 100% compatibility. To further reduce the likelihood of rater bias, each participant was rated by only one rater. Thus, some participants were rated by “Rater A” while others were rated by “Rater B.” This procedure was established to reduce rater bias. The raters were experimentally blind to the training received by each participant.

A further word on inter rater reliability may be useful. Establishing the same expectation from each of the items on the rater evaluation form was accomplished by having both raters evaluate the same 5 persons. This provided the opportunity to clarify the intent of each evaluation item and foster a common rating regimen early on in the experiment. However, since the same rater evaluated both the pre and post test for a pilot, the chance that a pilot might be rated by one rater in the pre or post test and differentially rated by the other rater in the corresponding post or pre test, respectively, was negated. As well, the raters evaluated about the same number of pilots so the impact of rater bias was further reduced.

The raters served as Air Traffic Control during the pre test and the posttest; as an annoying passenger during the flight that the pilot; as the simulated owner of an air taxi service who required accommodation and as the rater. The sequence of events and the choreography of the pre and posttest flights can be examined in Appendices 3, 12 and 16.

APPARATUS

Microsoft Flight Simulator (2002) was the software engine that drove both the traditionally instrumented Cessna 172 simulation device and the TAA instrumented Cirrus SR20 simulation device used in the study.

Figure 2 shows a photograph of the Elite system Cessna 172 simulation device used in the traditional aircraft proficiency test as explained in the Procedure section below.

Flight Simulator is a well-known and reliable product, particularly in the Microsoft Windows environment. A number of third party products support Microsoft as the industry standard so it is easily interfaced with other software architecture. Primary among these were the Go-Flight (radio, electrical, throttle, autopilot) and other flight instruments (yoke and rudder) and the Chelton TAA instrument. The data recording capability of Flight Simulator permits aircraft state data to be captured, including switch positions, dial and gauge position and pilot input to the flight instruments, on a second by second capture rate captured as data for the study.

An Elite flight simulation device, consisting of displays, yokes, instrument banks and rudders, similar to what might be experienced in a multi-engine aircraft was used to gauge pilot proficiency as explained in the procedures section below. The Elite system (Figure 2) was configured to be a Cessna 172 so for the multi engine Elite simulation device, only one display and engine instruments were used. The system utilized an overhead projection of the outside the window view on the wall directly in front of the pilot.

Figure 3 shows a photograph of the Chelton glass Instrument inside the Frasca 410 half cockpit used to simulate the Cirrus 20 aircraft used in the proficiency test, the Pre-test and the Posttest as explained in the Procedure section below.

The Rater sat behind the participant at the instructor station for the Elite system and had access to the Microsoft Flight Simulator Instructor software to control weather and other system failures without the participant aware of the changes.

The TAA aircraft simulation device used in the study utilized a Chelton MFD and PFD as well as several instruments from Go-Flight that directly interfaced with Flight Simulator set up to emulate a Cirrus SR20 aircraft. As shown in Figure 3, the aircraft environment and instruments were given a lot of detail to increase the realism of the tests. An overhead projector cast the outside the windscreen view on to a large projector screen directly in front of the pilot. The pilot sat inside a Frasca 410 half shell aircraft simulation device on an adjustable aircraft seat. The windows were blocked to restrict vision to an outside the windshield view only.

The Chelton was interfaced into the MS FS environment by the FSUIPC EFIS system developed by Peter Morton through a Gateway Interface Port. All peripheral devices were fed into USB connections to a Dell Computer hidden from the pilots view. All extraneous material was removed from the pilot's field of view to increase the feeling of realism. A video camera, seen in Figure 3, was used to videotape the pilot's manipulation of the Chelton system for later off screen analysis.

As with the Elite simulation device, the Rater and the Instructor had access to a nearby computer running Microsoft Flight Simulator Instructor Station that allowed easy access to the flight environment. Thus, the Rater or the Instructor could easily change weather or fail equipment through an Internet connection to Microsoft Flight Simulator running the Chelton and pilot displays.

PROCEDURE

Data collection lasted for 9 months beginning on 15 October, 2004 and was terminated 15 July, 2005. Participants were contacted two days before their pre-test session to remind them of their scheduled time in the study. The technician would greet the participant upon arrival at the TAAP lab. The technician would explain the procedures and caution the participant that he was not expected to know anything about TAA instrument flight and so they might find the pre and posttests difficult. The participants also were instructed not to discuss the conditions of their training with other potential participants, as it might be quite different for them. They were told that, for the purposes of the study, they were an air taxi service about to purchase a new SR-20 aircraft and the study was designed to determine the effectiveness of different instruction methods to learn TAA instrument. The participants then read and signed the informed consent document and began planning for the proficiency flights. The sequence of forms and events used in the study can be discerned from the Participant Data Form Checklist in Appendix 3. All the forms are listed in the Appendices.

The technical staff for the study assisted the Rater and the Instructor as well as ensured compliance with all the forms and questionnaires required of the study. They were available at all times during the pre and posttests, as well as the weekend instructions, to assist in any way. They were not naïve to the participant training, either NT, SBT or MBT but did not influence the rating or instruction. They also served to ensure that the Rater and for most of the experiment, the Instructors, were experimentally blind to the training condition of the participant.

The participants were required to fly two proficiency tests. These are explained in the forms used, Appendices 6 and 7. It was an opportunity to assess their instrument skills in a conventional aircraft, a Cessna 172, in the Elite simulation device and in the TAA simulation device. These tests were identical for both devices and consisted of a VFR flight out of Daytona Beach into a weather induced instrument flight. The tests were not meant to be difficult but to provide a quick estimate of the participant's ability as an IFR pilot. The order of proficiency tests was random in that about half the participants received the Cirrus simulation device first and the other half received the Cessna device first. In the design phase of the study, it was felt that proficiency tests would be a good way to screen out pilots who were overqualified in TAA and might have misrepresented their TAA experiences. None were found. It was also hoped that at some time in the future, pilot skills in the pretest might be used to predict success in FITS training. This may be the focus of a future report.

As indicated in Appendix 3, after the proficiency tests, the participants were asked to complete some workload and situational awareness forms. They then were told that for the tests to follow (they were not told it was a pre-test nor a baseline to prevent them from second guessing what the remainder of their participation would entail) they were an air taxi service picking up a new airplane at the manufacturer. They were then given the pre-test return route and began planning for the pre-test flight. Immediately upon completion of the flight planning, they were given what was called a “salesman’s brief”. This was the extent of the instrument training for the NT group and also was a shared experience with the MBT and the SBT groups. It was hoped that some small instruction on the TAA at the very beginning might make the results more meaningful on the pre test as the pilot might try to work the new system. This also would be similar to the kind of basic instruction they might receive after purchasing a new aircraft. The salesman’s brief highlighted the basic features of the Chelton PFD and MFD. It is reproduced in its entirety in Appendix 21. This brief was expected to provide all the participants a common background in training before they began more formal training or not, in the case of the NT group. The Technical person provided the salesman’s brief. After the salesman’s brief, they immediately got into the simulation device and began pre-flight procedures.

It should be pointed out that the Task 1&2 FITS team owes Cirrus, Inc a debt of gratitude for establishing a very important and successful TAA pilot training regimen for their new aircraft. They have been an important partner in the effort to standardize and improve the training of pilots new to TAA. In no way was the use of Cirrus in this study meant to imply anything other than that they are a leader in the manufacture of TAA and in efforts to effectively train the same and in fact, Cirrus uses the Avidyne Flight Max cockpit, not the Chelton system.

All participants got the proficiency tests, the salesman’s brief, the pre test and the posttest in addition to completing the questionnaires immediately after the pre and posttest as explained in the Participant Checklist, Appendix 3, below.

Figure 4 shows the procedure at a glance with the exception of the approximately two-hour pre and posttests. The pre and posttest sessions took about 3 hours each to complete and the Instructor training weekend sessions took about 4 hours each. The Rater sat behind the participant for the pre and posttests in such a way as to clearly see the manipulation of the aircraft instruments. Both they and the Instructors had access to the Flight Instructor computer to manipulate various events. The technician videotaped all the Chelton manipulations for subsequent offline evaluation.

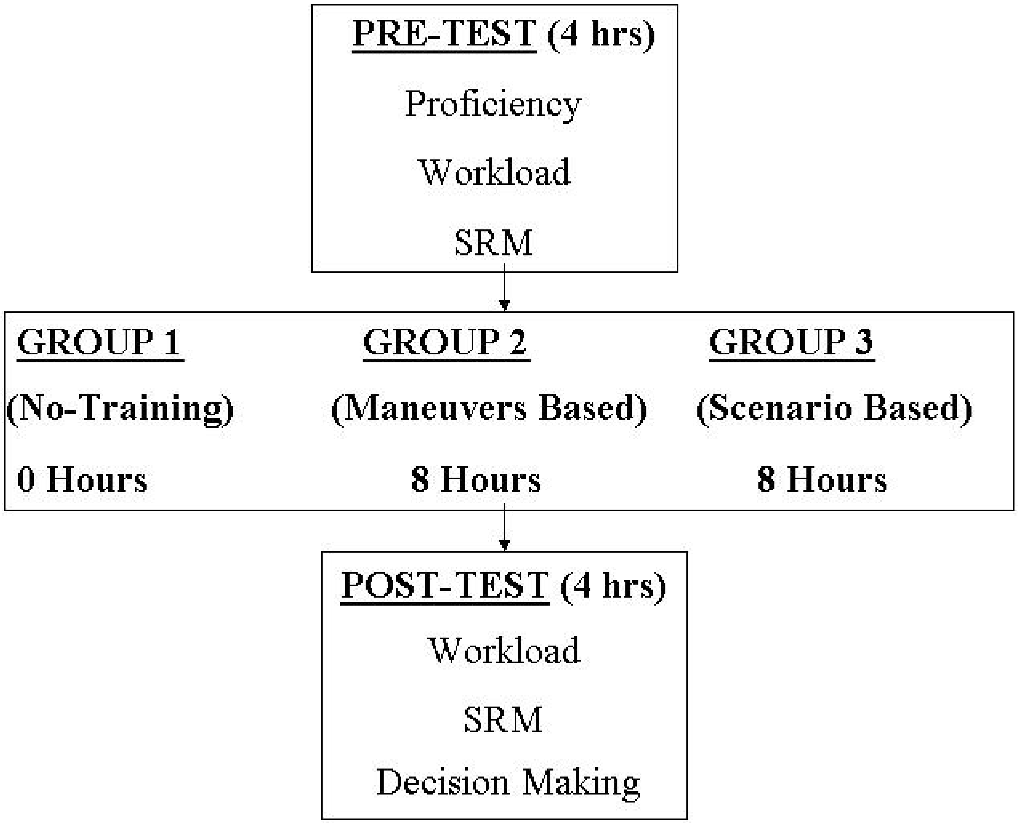

Figure 4. The procedure and some of the tests used in the pre test posttest experimental design. All subjects received the basic instruction in how to use the instrument during the "salesman's brief". Group MBT and SBT received an additional 8 hours of training each whereas group NT did not.

Most of the attitude surveys and other questionnaires used were new or did not have rigorous validation information to support them. Still, for most of the forms they had reasonable face validity to the technicians and subject matter experts to allow for their inclusion. The NASA Task Load Index (TLX) represents a noticeable exception, as it is an historical standard in the field for assessing workload (Colle and Reid, 1998). A second workload form was devised that attempted to break down workload into multiple cognitive resources of visual, cognitive, auditory and psychomotor (VCAP) abilities. After the posttest flight, similar questionnaires were completed as for the pre test with the exception of the Landa Workload questionnaire (developed by a graduate student at ERAU and included in the study to help in the validation process).

An effort was made to collect situational awareness data from the pilots in all groups. The Situational Awareness Rating Technique (SART) was included after the pre and posttest flights to see if situational awareness had improved as a result of training (Taylor, 1990).

Another goal of the study was to assess Judgment or Decision making skills. An FAA sponsored questionnaire has been developed for that purpose (Driskill, Weissmuller, Quebe, Hand, & Hunter, 1998) and was utilized in the study. According to Driskill (1998), studies at Illinois University using MIDIS (Microcomputer based Decision Simulator) evaluated pilot decisions making in this simulated environment. Results demonstrated that there is little relationship between pilot experience levels and performance decision-making tasks. Hence, they developed a questionnaire to get at the efficacy of pilot decision-making. Subject matter experts in aviation, particularly glass instrument flight and in training, modified the original Driskill questionnaire to address the unique kinds of decisions that must be made by TAA pilots. The modified Driskill questionnaire can be found in Appendix 17. These data were not analyzed in time for the final report and will be included in an amended report. The biggest reason the DM/JQ data were not analyzed is because at present, we don't know what the answers are. As can be seen from Appendix 17, the Decision Making/Judgment Questionnaire (DM/JQ) requires the pilot to indicate the likelihood of completing each alternative to a particular question (scenario). We have found differences in our expert opinion about the likelihood responses and so need to sample a larger pool of experts.

The videotape analysis was conducted off line by a Rater, experimentally blind to the participant and therefore to the participant's group. A copy of the rating form with the rating scale, is included in Appendix 22.

DATA ANALYSIS

All data were analyzed using the SPSS statistical evaluation software, version 12.0. The data were assumed to be non- parametric due to the non-interval scaling used in sampling pilot skills and attitudes and a Mann-Whitney U test was used to compare the MBT to the SBT in a two independent samples statistical test. This test is a well-known and robust test of the independence of different populations ranking.

All eight flight events from both the pre and the posttest rating tests, as explained under Performance Measures in the section above, were compared. In addition, every instance of the autopilot, which was on the checklist frequently throughout a flight, in many pre and posttest events were compiled to provide a look at the use of the autopilot as a result of training. The same approach concerned the GPS and the MFD instruments, independent of events in the pre and posttest, the use of either GPS or MFD were selected and a percent correct score obtained.

RESULTS

The tests in which SBT and MBT differed significantly ($p<0.05$) or produced a statistical result of ($p<0.10$) are graphed below, showing pre test and post test scores for both. A one tailed distribution was used for the alpha probability level since we hypothesized the direction of the SBT improvement over MBT. Although not significant, the NASA TLX data are presented to show a representative result that was at least visually in the predicted direction. In order to convey an idea of the central tendency of the results, the average and standard error of the SBT and MBT group means are graphed in each figure below. Although this is, strictly speaking, an incorrect way to show non-parametrically distributed data using a parametric measure of central dispersion, it was felt that this approach would convey the results better than simply presenting the non-parametric rank for each condition. For each graph then, the pre test and posttest scores are shown as an average plus the standard error of the mean. In the accompanying text, the Mann-Whitney U result is provided with the ranking results for each group.

The DM/JQ results were not shown due to the lack of having a standard with which to compare the results (i.e. expert data). The Higher Order Thinking Analysis (HOTA) data (Appendix 11) also were an experimental measure and are not shown since the results were not significant.

Those flight events (see Methods section above) that were determined to represent independent populations between the MBT and the SBT for the Rater derived posttest scores in that a statistically significant difference was found were the following:

Autopilot Use (see explanation above)
Event 2: Pre-Flight Preparation
Event 5: Re-route
Event 7: Approach
Event 8: Missed Approach

Those posttest flight events in which equivalence of the two population distributions were found (non significance) were:

GPS use (see explanation above)
Event 4: Take-off and Departure
MFD use
Event 1: Flight Planning
Event 3: Pre-Take-off
Event 6: En-route
Videotape analysis

None of the questionnaire data demonstrated statistical significance, that is, subjectively the MBT and SBT pilots seemed to represent the same posttest population.

The 5 remaining pilot subjective questionnaire data, the Self Efficacy Questionnaire – Single Pilot Resource Management (SEQ-SRM) questionnaire, the SART, HOTA and NASA Task Load Index (TLX) questionnaires and a measure of Visual, Cognitive, Auditory and Psychomotor workload (VCAP) did not indicate a statistical difference. These represent the extent of the data evaluated for this report.

Autopilot Use
The autopilot scores were obtained by compiling all the criteria in each flight event during the pre and posttests that required an autopilot use on the Rater checklist (see Appendices 12 and 16). The posttest difference between the MBT (mean rank 10.47) and SBT (mean rank 18.42) was represented by a Mann-Whitney U value of 37 and had a p value of 0.004. Figure 5 shows that the use of autopilot was performed correctly more often by the SBT group compared to the MBT group.

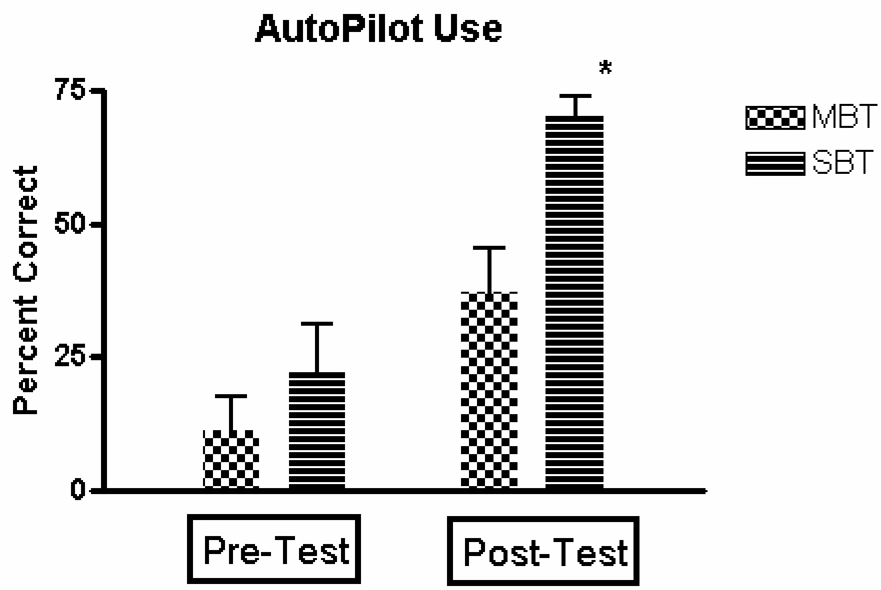

Figure 5. Autopilot useage in Pre- and Post tests for Maneuvers Based Training (MBT) and the Scenario Based Training (SBT) Groups. For all the graphs to follow the data are shown as a mean percent correct with the standard error of the mean shown.

Event 2: Pre-Flight Preparation

The pre-flight preparation event occurred in Vero Beach, FL for the pre test and from Daytona Beach, FL for the posttest. In spite of the differences in location and ATIS information, the requirements for the pre and posttests were the same. As with all the flight events used to score piloting and navigation, the percent correct of the criteria for each flight event were used as the event score (see Methods). Figure 6 shows that the SBT (mean rank 18.83) correctly performed the items on the preparation checklist more than the MBT (mean rank 10.13) Group (U= 32; p=0.002).

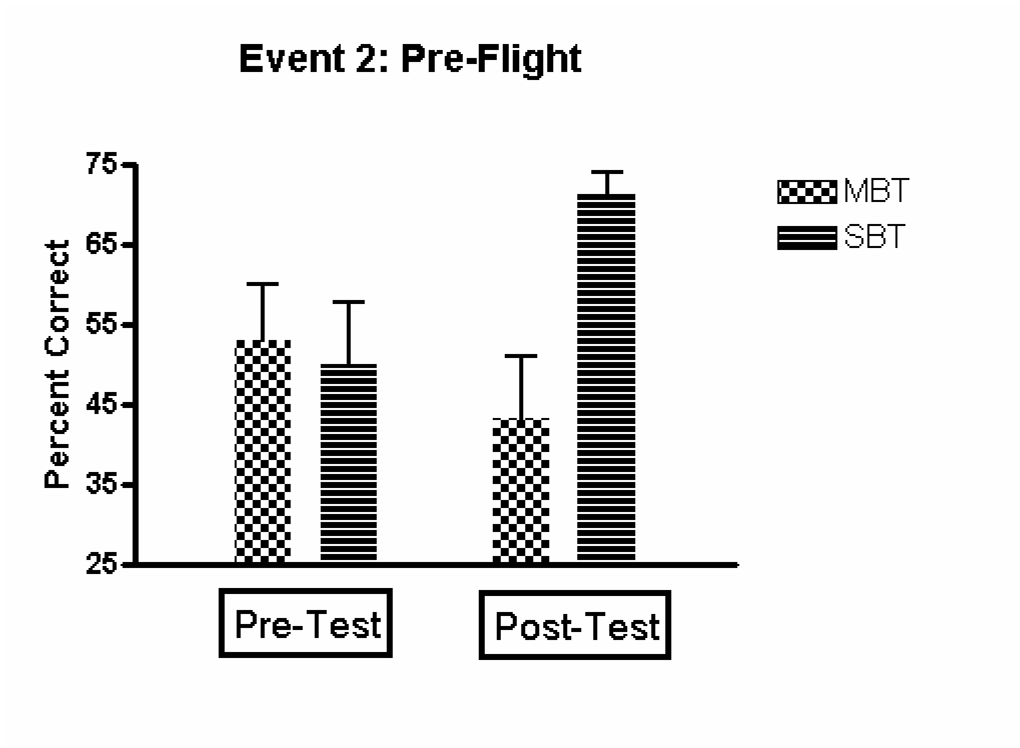

Figure 6. Pre-flight preparation for MBT and SBT pilots. This flight event showed SBT to be significantly better than MBT.

Event 5: Re-route

During the flight, the weather produced demanding instrument flight for both pre and posttests. The SBT pilots (mean rank 17.7) did better than the MBT pilots (mean rank of 11.47) as shown in Figure 7 (U=52.0; p=0.03). It is important to note that both MBT and SBT pilots showed a decrease in this measure, the SBT pilots less so, when considering the average percent correct score.

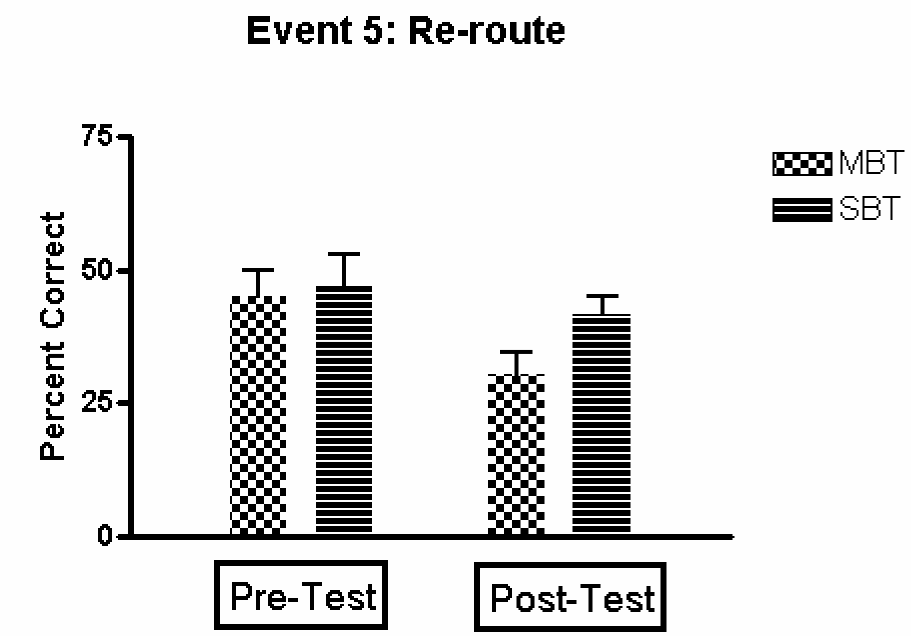

Figure 7. Re-Route Event for MBT and SBT pilots during flight. This flight event showed that SBT tended to be different from MBT although both groups decreased in the posttest compared to the pre-test.

Event 7: Approach

The Approach phase of flight was performed better during the posttest by the SBT group (mean rank of 17.7) compared to the MBT group (mean rank of 11.07) as shown in Figure 8 (U=46; p=0.015)

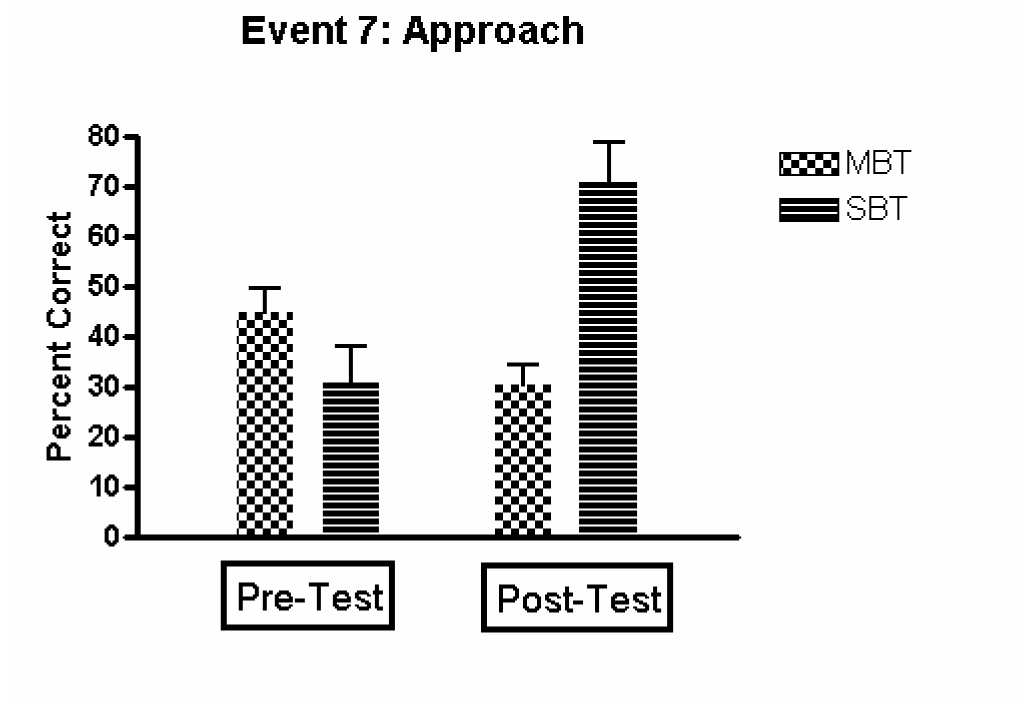

Figure 8. Approach Phase of flight for MBT compared to SBT. This flight event showed SBT to be significantly better than MBT.

Event 8: Missed Approach

The piloting and navigation responses to a missed approach during the posttest was better for the SBT group (mean rank of 18.08) compared to the MBT group (mean rank of 10.73) as shown by the graph in Figure 9 ($U=41$; $p=0.008$).

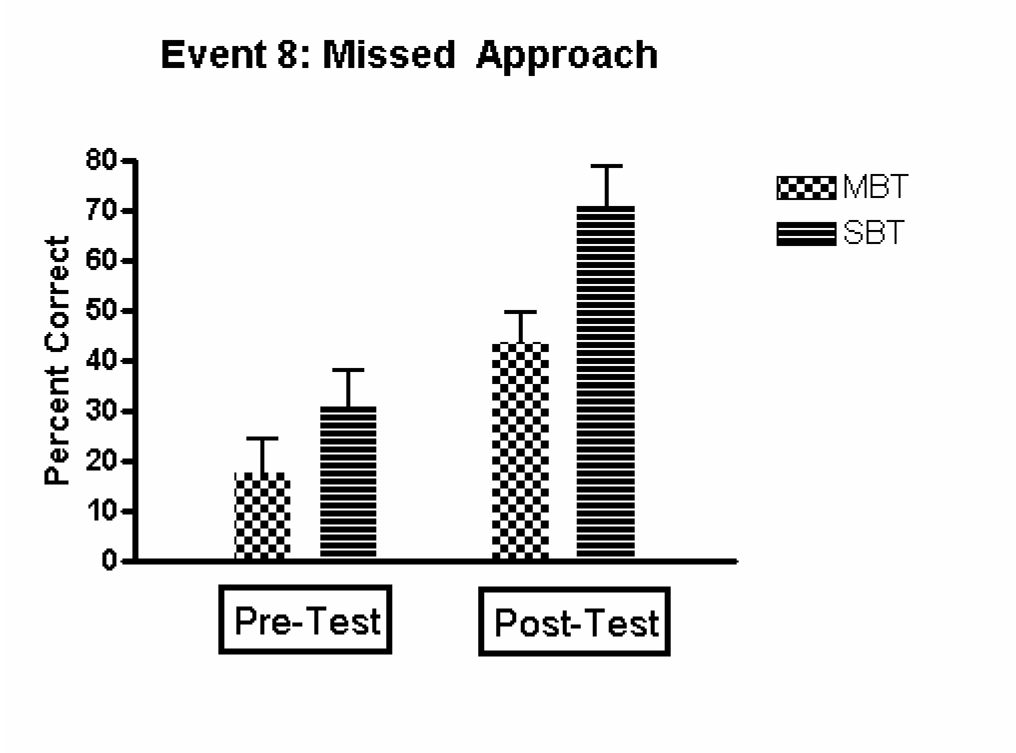

Figure 9. Missed approach event for MBT compared to SBT. This flight event showed SBT to be significantly better than MBT.

The remaining results indicated a small but not significant result towards improvement by the SBT group over the MBT group that might have been achieved with a larger sample size.

GPS use

The flight event criteria for GPS use occurred throughout the pre and posttests. These were compiled into a single event and is shown as percent correct usage in Figure 10. The posttest data were suggestive but not significant for SBT (mean rank of16.0) over the usage registered for the MBT (mean rank of 12.4). The Mann –Whitney U for this result was 66 with an alpha level of 0.1.

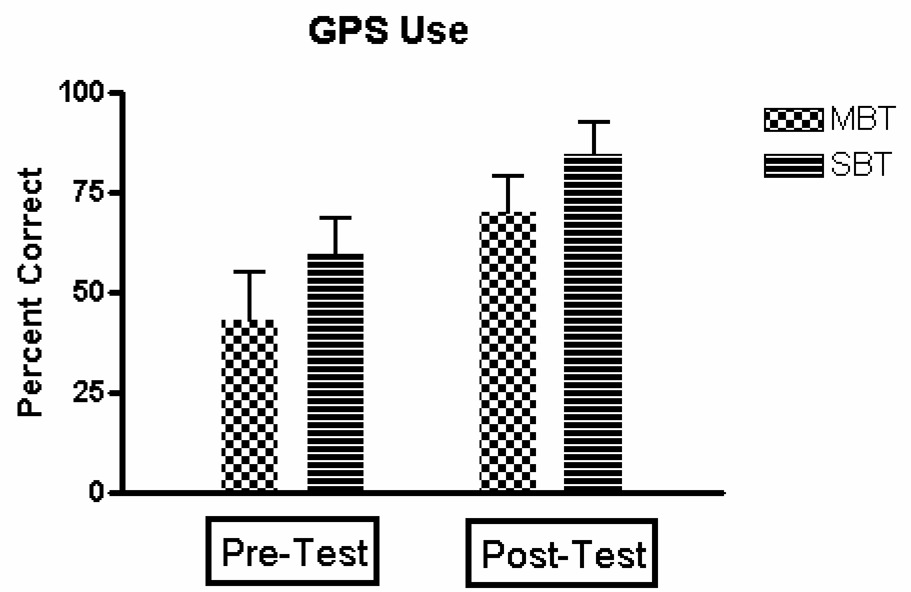

Figure 10. GPS use across all flight events for the pre and posttest MBT compared to SBT pilots. This flight event showed that SBT tended to be better than MBT.

Event 4: Take-off and Departure

During the take-off and departure flight event for the posttest, there was a tendency for the SBT pilots (mean rank 16.67) to do better than the MBT pilots (mean rank of 11.87) as shown in Figure 11 ($U=58.0$; $p=0.05$).

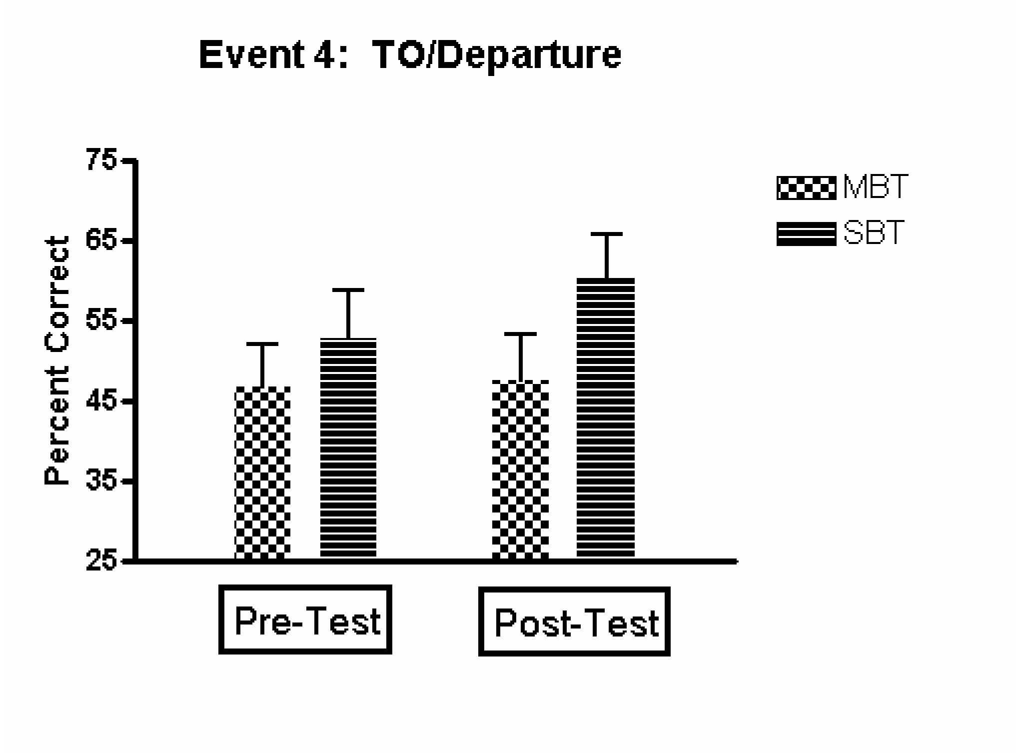

Figure 11. Take-off / departure event for the pre and posttest MBT compared to SBT pilots. This flight event showed that SBT tended to be better than MBT.

Self-Efficacy TAA

The posttest scores of self efficacy for Technically Advanced Aircraft (Appendix 9) tended to show an increase for the SBT pilots (mean rank of 13.81) compared to the MBT pilots (mean rank of 9.7) as shown in Figure 12 (U=41.5;p=0.07).

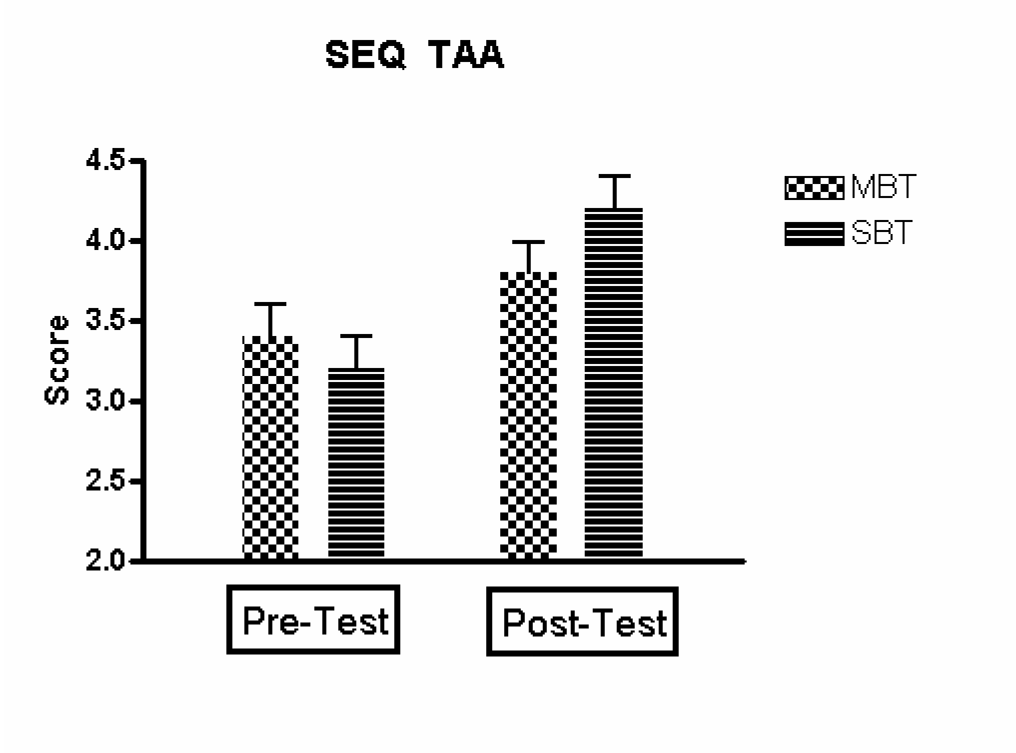

Figure 12. Subjective self-efficacy for the pre and posttest comparison of MBT and SBT.

TLX

The NASA TLX results were not significantly different between MBT and SBT. However, these results were shown to indicate that, like the other non-significant subjective data, that visually, non-significant differences can be shown in favor of the SBT, in this case, a reduction in workload.

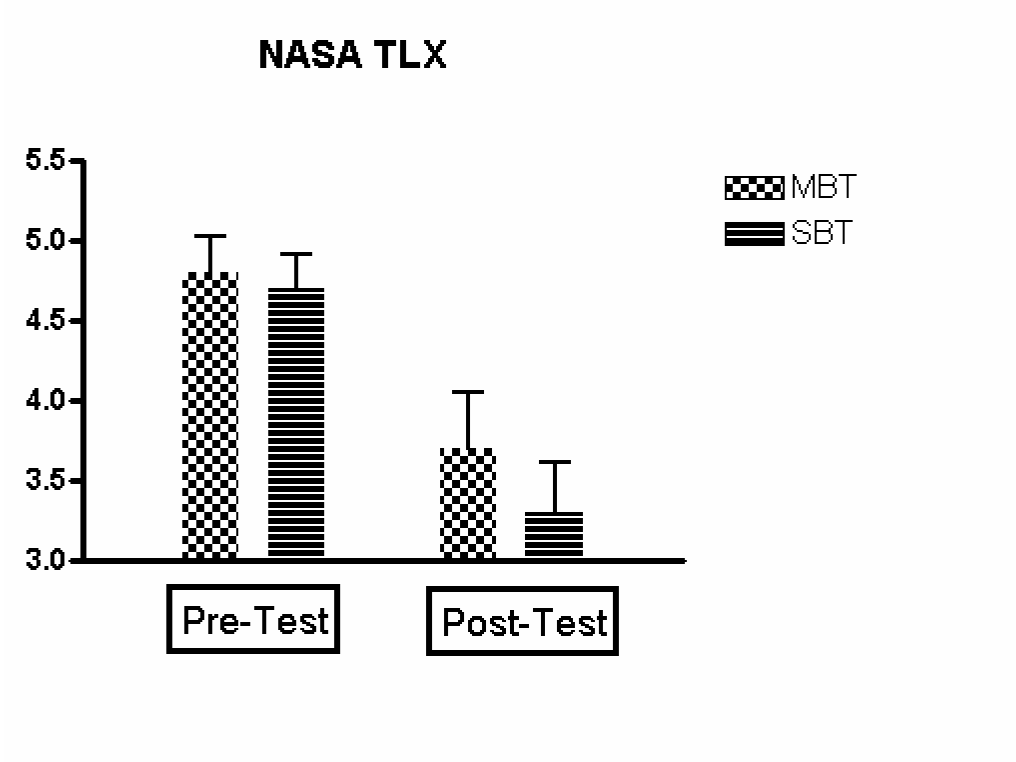

Figure 13. The Subjective NASA TLX workload scale for the pre and posttest comparison of MBT and SBT pilots. SBT was visually better than MBT, not statistically. This result was similar to the other non-significant measures.

The results of a survey conducted by an aircraft manufacturer in response to the FITS training that they provided to the pilots who had recently purchased one of their new aircraft is presented in Appendix 25. These data indicate that transitioning FITS SBT training to voluntary pilot education was well accepted and deemed useful by the respondents.

DISCUSSION

The study was designed to provide a fair comparison between traditional (maneuvers based) training, the standard for decades in instrument flight instruction and scenario-based training, a new approach to instrument flight training that draws on that distinguished past as well as on new concepts in education, such as self-evaluation. The results argue that scenario-based training is better than task oriented or maneuvers based training on most measures of piloting and navigation proficiency as rated by experimentally blind expert raters. On the measures where statistical significance was not found to indicate SBT was better than MBT, SBT was found to at least show parity with MBT. Further, the subjective attitude and workload questionnaires indicated that there was functional equality between the two training techniques. Where statistical significance was not achieved, it is likely that an increase in sample size for the two groups might have produced a difference between the two populations. Visual inspection of the data graphs would indicate that increasing the sample size would further support

SBT over MBT. In short, SBT pilots performed better than MBT pilots on most measures. The attitudes of the SBT pilots were in the right direction in most of the attitudinal and workload measures but were not statistically different. This suggests that pilots may demonstrate a measurably improved performance without necessarily being aware of it.

The goals of Task 1&2 were to address 5 points regarding FITS training or SBT. These are reiterated here:

- Improvement in judgment and decision making skills
- Improvement in automation Management skills
- Improvement in situational awareness
- Identify differences in FITS and Non FITS trained Pilot skills
- Identify specific weaknesses and strengths of the FITS program.

With regards to measuring improvements in judgment and decision making, the DM/JQ questionnaire data (Appendix 17) were not evaluated since not enough experts in TAA were found at the time of this writing that could take the questionnaire and provide us with a standard with which we could compare the participant's responses. It is very difficult to measure the psychological construct of Judgment and Decision Making but the test put out by the FAA (Driskill et al., 1988) is a good start. We believe that our subject matter experts have improved on this early test and related it specifically to SBT (Appendix 17). These data will form the basis of an amended report when those data are analyzed and compared to subject matter expert results. We hope in a future project to compare expert TAA pilots with novice TAA pilots to determine what kinds of decisions both make. In that way, it may be possible to identify someone's piloting decision-making ability as they progress from novice to expert. The results of the Rater data for the posttest indicate that the performance of the SBT group was, for the most part, improved over MBT. We have taken this to mean that SBT pilots made better decisions overall in the tests of piloting performance and we have met this goal.

The changes in Pre and Posttest scores are evident from the data graphs shown. In most cases, both SBT and MBT produced a statistical improvement from the pre to the posttest. This suggests that both training methods were effective and registered an improvement on the evaluation scores. There were many ways the data could have been presented and evaluated, the difference score between pre and post test results for each pilot for example, may have 'normalized' the data relative to the percent change on the post test. However, the authors felt it important to present the pre-test data graphically to demonstrate that although slightly different in appearance, the pre-test scores were not significantly different between the MBT and the SBT groups. The differences highlighted by this graphical presentation are clear a.) between the pre and the post test scores for both MBT and SBT, demonstrating efficacy in training of either sort, MBT or SBT and b.) between MBT and SBT in the post test scores demonstrating in several cases that SBT produced better scores. The differences between the pre and posttests were not the focus of this final report though and will not be discussed further.
The videotape assessment was not as useful as was originally hoped. The camera was too far away from the Chelton to read the screen and the fine adjustments made by the

pilot. Certainly major adjustments and screen changes were clear from the camera position immediately behind and to the right of the pilot. The analysis was planned to focus on the number of button presses and knob adjustments which we presumed would be very high in frequency in the pre test and less in frequency for the post test. This was the case for the most part but the variability of this adjustment frequency was too great to admit any statistical delineation between the training methods. Also, it became too labor intensive to find the exact spot on the video tape that would allow comparisons between events on the pre and the post test. That efforts to mark these events were not entirely successful was not discovered until after the data had been collected. Still with the voice captured and major events in the scenario visible on the videotape, this approach may well be useful in future analyses. The authors believe a better approach might be to have a subject matter experts fly both pre and post tests and evaluate responses of the participant pilots relative to the experts. What was clear from the videotape analysis was that this metric did not have the resolution needed to see any differences between MBT and SBT and so the emphasis on this report was on the more sensitive rater evaluation form.

The data from the autopilot comparisons of SBT and MBT (Figure 5) during the posttest, argue that the second goal of Task 1&2 was met. By focusing the analysis on those flight incidents in which autopilot use was required, SBT pilots were better than MBT in automation management. They had more time to focus on other events during flight and their time was better managed because the autopilot was engaged.

The third goal of the Task 1&2 team was to demonstrate improvements in situational awareness. This is another very difficult concept to address but literature suggests that it can be measured and we hope that our SART questionnaire (Appendix 13) was an adequate measure. While not significant, the data indicate that situational awareness, as measured by SART, was improved in both the SBT and the MBT from a visual inspection of the graph. As situational awareness is another psychological construct that is amorphous and hard to measure, the lack of a sensitive assessment tool may explain the failure to find a difference in SA between the two groups.

The fourth goal set before Task 1&2 concerned the differences between the MBT pilot and the SBT pilot. We believe our results indicate where those differences exist, in piloting performance, in attitude and in subjective impressions of workload. SBT pilots tended to demonstrate better scores on these measures than MBT pilots. Future research should exploit these findings to further refine the differences. One recommendation for future research would be to evaluate the visual scanning technique that results from the two training techniques. It may be that SBT pilots have a different visual scan and this might improve MBT pilots if it could be trained. This research recommendation would be important on other grounds. There is very little empirical data relating visual scan technique to piloting performance. A recommended project would be to use the current results to design a study that evaluated eye tracking during simulated flight and compare the results with piloting performance; both in traditional and TAA paradigms.

The final goal of Task 1&2 was to identify specific strengths and weaknesses of the FITS program. This goal seems far beyond the scope of the research data presented here which was one project in the FITS program. However, the FITS training methodology seems, on the face of our results, to be a regimen that is sound from an empirical perspective. In addition, SBT has much face validity and has been accepted well in other cognitive domains, particularly where computerized technology is used, such as medicine and nuclear fuel source industries. The strengths of SBT then, seem to lie in the results reported here that, as good as traditional training methods are, the MBT, the SBT method of teaching glass instrument flight seems to be better. This is the first case of which we are aware that SBT was compared directly with MBT and the results support SBT. This is not to say that MBT is a bad training regimen. Indeed it has been used for many decades to successfully train instrument pilots. SBT seems better in this case.

Still, there are several questions posed by this research that should form the basis of future research to support the FITS program. Some of the weaknesses of the training methodology arise not from SBT but from the inadequacy of the measures used. The Decision Making questionnaire for example, would be an important tool if it could identify expert from novices easily. Similarly, the visual scan technique could be an important adjunct to FITS Training if an ‘expert’ scan could be identified. The Task 1&2 team has proposed research as a future projects that could identify ways to improve these measures. Another weakness of the Task 1&2 project is that it has yet to evaluate the duration of SBT training, how long the improved results last. It is important to know how long the SBT training lasts, how long before a pilot looses the improvements brought about by SBT. This could provide an important distinction between SBT and MBT and also should be considered for future funding. The large participant group, many of whom still remain on campus, would provide an immediate answer to this question if re-tested in a longitudinal manner, at the 6 months and one-year intervals for example.

One large issue for the FITS training that seems important to address before the FITS program is fully vetted and absolutely should be the focus of future research relates to scenario development in scenario-based training. Effective scenarios need to be identified that are necessary and sufficient to bring about the improvements in piloting ability demonstrated here. Perhaps not all scenarios are effective; perhaps some scenarios are more effective. These results would also provide an answer to the question of how long training needed to be to still impart the benefits of FITS training. We hope that the results of our study will fuel the interest in further support for addressing these questions.

SBT emphasizes whole task training, using realistic scenarios, tightly coupled learning objectives, performance measures, and feedback. This report contrasts SBT with the maneuvers based approach to training instrument flight in TAA. The MBT approach has served aviation well for many decades. However, recent and radical advances in avionics that improves safety have been rapidly adopted by the aviation industry. The basic premise of this report is that it is the right time for advances to be made in pilot training as well, particularly for the new avionics embodied by the TAA. This is the goal of

FITS. The results of a survey conducted following a FITS voluntary training class for pilots purchasing a TAA indicates that FITS may be well accepted in the operational settings (Appendix 25). This Final Report for Tasks 1&2 suggests that the goal of FITS has been met, at least as measured in the laboratory. The concept of FITS seems sound in improving piloting and navigation proficiency in TAA. It remains up to the academic, industry and FAA partners in FITS to make it work in application.

REFERENCES

Cannon-Bowers, J.A., Burns, J.J., Salas, E., & Pruitt, J.S. (1998). Advanced technology in scenario-based training. In J.A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 365-374). Washington, D.C.: APA.

Chelton Flight Systems. (n.d.) *Flightlogic features/benefits.* Retrieved January 31, 2005, from http://www.cheltonflightsystems.com/Prod_cert_features.html

Colle, H. A., & Reid, G. B. (1998). Context effects in subjective mental workload ratings, *Human Factors, 40(4)*, 591-600.

Driskill, W.E., Weissmuller, J.J., Quebe, J., Hand, D.K. and Hunter, D.R. (1998). Evaluating the decision-making skills of general aviation pilots. *FAA Office of Aviation Medicine;* Washington, DC.

Dwyer, J.D., Oser, R.L., Salas, E., & Fowlkes J.E. (1999). Performance measurement in distributed environments: Initial results and implications for training. *Military Psychology, 11*(2), 189-215.

FAA/Industry Training Standards. (2004). FITS Curriculum Recognition Criteria. Retrieved January 28, 2005, from http://learn.aero.und.edu/pages.asp?PageID=13814

Goldstein. I. L. (1993). Training in organizations. Pacific Grove, CA: Brooks-Cole.

Kneebone, R.L., Nestel, D., Moorthy, K., Taylor, P., Bann, S., Munz, Y., et al. (2003). Learning the skills of flexible sigmoidscopy – the wider perspective. *Medical Education, 37*(1), 50-58.

Loftin, R.B., Wang, L., & Baffes, P. (1989). Intelligent scenario generation for simulation-based training. In *AIAA Computers in Aerospace Conference, 7th, Monterey, CA* (pp. 581-588). Washington, DC: American Institute of Aeronautics and Astronautics.

Lowry, K.D. (2000). United States probation/pretrial officers' concerns about victimization and officer safety training. *Federal Probation, 64*(1), 51-55.

Mulder, J.A., van Paassen, M.M., & Mulder, M. (2004). Perspective guidance displays show the way ahead. *Journal of Aerospace Computing, Information, and Communication, 1*(11), 428-431.

NASA Facts Online. (2000). *Glass Cockpit Fact Sheet* (FS-2000-06-43-LaRC). Retrieved January 31, 2005, from http://oea.larc.nasa.gov/PAIS/Glasscockpit.html

Oser, R.L., Cannon-Bowers, J.A., Salas, E., & Dywer, D.J. (1999a). Enhancing human performance in technology rich environments: guidelines for scenario-based training. In E. Salas (Ed.), *Human/Technology Interactions in Complex Systems* (pp. 175-202). US: Elsevier Science/JAI Press.

Oser, R.L., Cannon-Bowers, J.A., Sals, E. and Dwyer, D.J. Enhancing human performance in technology-rich environments. Guidelines for scenario-based training., In Human Technology interaction in complex systems, Edwardo Salas (Ed.) Vol. 9., pp 175-202. Elsevier Science Press.

Oser, R.L., Gualtieri, J.A., Cannon-Bowers, J.A., & Salas, E. (1999b). Training team problem solving skills: an event-based approach. *Computers in Human Behavior, 1*5, 441-462.

Salas, E. and Cannon-Bowers, J.A. (2000) The anatomy of team training. In S. Tobias and J.D. Fletcher (Eds.), Training and Retraining: A handbook for business, government and the military (pp 312-335), New York, Macmillian.

Scharr, T.M. (2002). Interactive video training for firearms safety. *Federal Probation, 65*(2), 45-51.

Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. In: AGARD Conference Proceedings, no. 478, Situational awareness in aerospace operations. Aerospace Medical Panel Symposium, Copenhagen, October 2-6, 1989.