Initial development of a metric to describe the level of safety associated with piloting an aircraft with Synthetic Vision Systems (SVS) displays

Anthony P. Bartolone[a], Louis J. Glaab[a], Monica F. Hughes[a], Russell V. Parrish[b]
[a]NASA Langley Research Center, 24 West Taylor St., Hampton, VA, USA 23681;
[b]Raytheon-ConITS, 130 Research Drive, Hampton, VA USA 23666

ABSTRACT

Synthetic Vision Systems (SVS) displays provide pilots with a continuous view of terrain combined with integrated guidance symbology in an effort to increase situation awareness (SA) and decrease workload during operations in Instrument Meteorological Conditions (IMC). It is hypothesized that SVS displays can replicate the safety and operational flexibility of flight in Visual Meteorological Conditions (VMC), regardless of actual out-the-window (OTW) visibility or time of day. Throughout the course of recent SVS research, significant progress has been made towards evolving SVS displays as well as demonstrating their ability to increase SA compared to conventional avionics in a variety of conditions. While a substantial amount of data has been accumulated demonstrating the capabilities of SVS displays, the ability of SVS to replicate the safety and operational flexibility of VMC flight performance in all visibility conditions is unknown to any specific degree. In order to more fully quantify the relationship of flight operations in IMC with SVS displays to conventional operations conducted in VMC, a fundamental comparison to current day general aviation (GA) flight instruments was warranted. Such a comparison could begin to establish the extent to which SVS display concepts are capable of maintaining an “equivalent level of safety” with the round dials they could one day replace, for both current and future operations. A combination of subjective and objective data measures were used in this research to quantify the relationship between selected components of safety that are associated with flying an approach. Four information display methods ranging from a “round dials” baseline through a fully integrated SVS package that includes terrain, pathway based guidance, and a strategic navigation display, were investigated in this high fidelity simulation experiment. In addition, a broad spectrum of pilots, representative of the GA population, were employed for testing in an attempt to enable greater application of the results and determine if “equivalent levels of safety” are achievable through the incorporation of SVS technology regardless of a pilot’s flight experience. This research was conducted under the Aviation Safety and Security Program’s (AvSSP) GA Element of the SVS Project at NASA Langley Research Center in Hampton, Virginia.

Keywords: SVS, synthetic vision, equivalent safety, general aviation, head down displays, highway-in-the-sky, HITS

1. INTRODUCTION

The recent advent of affordable electronic displays, with dimensions and other characteristics suitable for installation in small airplanes, offers general aviation avionics and airplane manufacturers more flexibility in presentation and placement of cockpit instruments and system controls than ever before possible (GAMA, 2004). These new technologies, which include SVS displays, can provide dramatic improvements in the capabilities of 14 CFR Part 23 certificated airplanes flown by a single pilot in IMC, according to the General Aviation Manufacturers Association (GAMA). Even though the introduction of electronic displays and/or systems in GA aircraft has been predicted for some time, the Federal Aviation Administration (FAA) has not clearly defined the certification methods or processes that would be used to qualify said systems for installation in single pilot operation aircraft for today’s operations. Therefore, the major avionics manufacturers are developing SVS, or other advanced displays for GA aircraft, with limited ability to accurately predict the certification process to an acceptable level of financial risk. The inability to forecast the level of financial risk associated with certification severely challenges the development and implementation of new avionics technologies like SVS. The FAA, and the GAMA have recognized the lack of guidance available to the industry it represents, and published, respectively, *GAMA Publication #12: Recommended Practices and Guidelines for an Integrated Cockpit/Flightdeck in a 14 CFR Part 23 Airplane* and *FAA AC 23-23* in an effort to ensure that safe operations are maintained as revolutionary electronic displays begin their introduction in the GA fleet. However, it is not the intention of the GAMA publication to require manufacturers to display or even offer every function, system, or control in the same position or method on all products but rather to ensure that all manufacturers provide uncluttered, easily interpreted control and display features which provide the basic information needed by a pilot to aviate, navigate, and communicate effectively. This achievement means generating and adhering to uniform basic control layouts, color coding, font sizing, range scaling, etc, which is what the GAMA publication attempts to. While the *GAMA Publication #12* and *FAA AC-23-23* serve well to standardize the electronic displays that are becoming increasingly more commonplace, their guidance falls short for revolutionary new systems, such as SVS. Similarly, the certification of new systems and displays, especially innovative new systems like SVS, requires the development of consistent certification criteria by the FAA which will necessitate adequate metrics to insure an "equivalent level of safety" is maintained with a candidate system or display as compared to the conventional accepted standard for traditional operating procedures. However, although adding to an already complex certification system, new avionics like SVS can potentially lead to new types of operational concepts. For example, initial terrain awareness and warning systems (TAWS) only issued "pull up" commands. Recent advances in terrain database avionics systems now include "avoid terrain" commands reflecting increased confidence in these types of system.

In the realm of cockpit display research, the task of flying an approach, regardless of a pilot's experience or maneuver type, revolves around the availability of flight critical information and the processes of acquiring and interpreting that information to form a mental model of the situation (Endsley, 1999). Subsequently, the readability of advanced cockpit displays plays an integral role in maintaining or improving safety. The construct of information display readability implies both a cost and benefit to the pilot when flying an approach. The cost of readability is mental workload and the benefit is situation awareness. Therefore, when designing and testing advanced cockpit displays, it is the responsibility of the human factors researcher(s) to optimize the information display such that the cost or workload required to acquire and interpret flight critical information is low and the benefit or situation awareness gained from the flight critical information is high. Subsequently, in order to properly evaluate a specific condition, it is necessary to measure not only the readability of the display, but also the intuitiveness of the information presented and the workload expended by the pilot to gather and process the information, along with the situation awareness the pilot was able to achieve as a result.

The ability to comprehensively maintain the relative level of safety is essential in order to facilitate certification and ultimate effective use of new avionics equipment. The development of a metric to describe the level of safety associated with piloting an aircraft with Synthetic Vision Systems displays is an integral part of the SVS-GA Equivalent Safety (SVS-ES) experiment research. Within the realm of synthetic vision, a significant amount of research to date has been focused on demonstrating the ability of SVS displays to make IMC operations resemble those conducted in VMC in terms of safety and operational flexibility. The results of the SVS-GA Terrain Portrayal for Head Down Displays (TP-HDD) simulation and flight test and the Symbology Development for Head Down Display (SD-HDD) simulation experiments (Glaab & Hughes, 2003; Takallu et al., 2004; Wong et al., 2004), as well as other SVS research conducted by NASA (e.g. Prinzel & Arthur et al., 2004; Kramer et al, 2003), academia, and industry, all indicate that a pilot's situation awareness (SA) can be substantially improved when flying in IMC by providing a SVS terrain background on a primary flight display (PFD) or even Head-Up Displays (HUDs) (Prinzel & Comstock et al, 2004). Furthermore, overlaying the SVS terrain with a highway-in-the-sky (HITS), or other pathway based guidance symbology, results in very precise aircraft maneuvering without causing a significant increase in the workload associated with flying an approach, executing a missed-approach, or flying enroute maneuvers (Glaab & Hughes, 2003; Kramer et al, 2003; Takallu et al., 2004; Wong et al., 2004; Prinzel & Arthur et al, 2004). Even though these piloted simulations and flight tests have shown better SA and path precision is achievable with SVS displays without causing an increase in workload, none of the previous SVS research attempted to fully capture the significance of SVS displays in terms of their contribution to safety or operational benefits. In order to facilitate such a connection, a fundamental comparison between SVS displays and current day general aviation avionics instrumentation was required. Such a comparison was the focus of the SVS-ES experiment.

2. METHODOLOGY

2.1 Evaluation Pilots

A broad spectrum of pilots, representative of the GA population, was employed for the SVS-ES experiment in an attempt to more fully quantify, and enable a broader applicability, of the results from this experiment. Twenty-five evaluation pilots (EP) with a minimum of a current valid FAA Class III Medical Certificate were recruited to participate in this experiment. The EP's were categorized by their experience level. The resulting pilot categories, along with descriptive statistics, are provided below (Table 1).

Pilot Category	N	Mean Age (yrs)	Mean Experience (hrs)
VFR (<400 hrs)	8	36	178
IFR (<1000 hrs)	8	42	465
H-IFR (1000+ hrs)	9	49	5,444
Total	25	43	2,291

Table 1. Evaluation Pilot (EP) Data

2.2 Display Concepts

Four research display concepts (DC) were developed for incorporation into the IFD for the SVS-ES experiment. The baseline round dials (BRD) displays were traditional "steam gauge" style GA aircraft flight instruments, which included airspeed, altimeter, attitude, vertical speed, heading, turn coordinator, localizer-glide slope, and combined manifold pressure and prop RPM round dials (BRD; Figure 1 – DC 1). The second display concept was a generic 6" diagonal PFD with a single cue flight director (SCFD; Figure 1 – DC 2) coupled with the 6" diagonal strategic navigation display (Figure 1 – ND). The third display concept was also a 6" diagonal PFD which included the NASA crow foot tunnel with tadpole highway-in-the-sky (HITS PFD; Figure 1 – DC 3), again combined with the ND. The fourth and final display concept for the SVS-ES experiment was the fully integrated SVS package, which included an elevation based generic (EBG) terrain background on the PFD overlaid with the NASA crow foot tunnel with tadpole HITS (SVS+HITS; Figure 1 – DC 4), coupled with the ND. The NASA crow foot with tadpole tunnel HITS used in this research is a derivative of the "minimal" tunnel concept developed for the commercial and business aircraft segment of SVS research at NASA Langley (Prinzel et al., 2004). The combination HITS and terrain was adopted for GA application for the SD-HDD series of experiments performed previously by the SVS-GA team (Takallu et al, 2004; Wong et al, 2004). Also, Due to increased display space availability on the RD when portraying DC's 2, 3, and 4, RPM and manifold pressure were presented on separate gauges directly below the ND.

2.3 Simulation Facility

Conceived as a coordinated simulation and flight test experiment, the SVS-ES flight portion will employ the NASA LaRC Cessna 206 Stationaire (C-206) research aircraft. The SVS-ES simulation experiment was conducted at NASA Langley Research Center (LaRC) using the Integrated Flight Deck (IFD) transport category fixed-base high-fidelity flight simulator. The IFD, nominally a Boeing 757 cockpit, was adapted for the SVS-ES experiment to take advantage of its excellent visual, tactile, and audio capabilities. A basic six degree of freedom non-linear simulation model of the C-206, a derivative of Roskams' Cessna 172 model, was created for this experiment (Roskam, 1979). A control force model was created using wind-tunnel data (Greer, 1973; NASA LaRC, 1980; Riley, 1994) combined with piloted evaluations by NASA LaRC research test pilots during simulation development and checkout. The resulting control force model was used to provide breakout and dynamic force-feel cues to the pilot via the McFadden hydraulic control loader. Dynamic aerodynamic data was obtained from the instrumented NASA LaRC C-206 flight test aircraft and incorporated into the simulation model. In addition, to further minimize any effects of conducting a GA simulation experiment in a transport cockpit, extraneous displays and gauges were covered, or turned off, during data collection. An electronic Research Display (RD) was installed on the instrument panel directly in front of the left seat and control yoke. The RD was composed of two 10.4" LCD displays and was capable of displaying two separate digital displays, side-by-side, and was the only source of critical flight information provided to the EP's (Figure 2). The collimated out-the-window (OTW) scene was produced by an Evans and Sutherland ESIG 4530 graphics system. The resulting OTW scene provided approximately 200 degrees horizontal by 40 degrees vertical field of view at 26 pixels per degree.

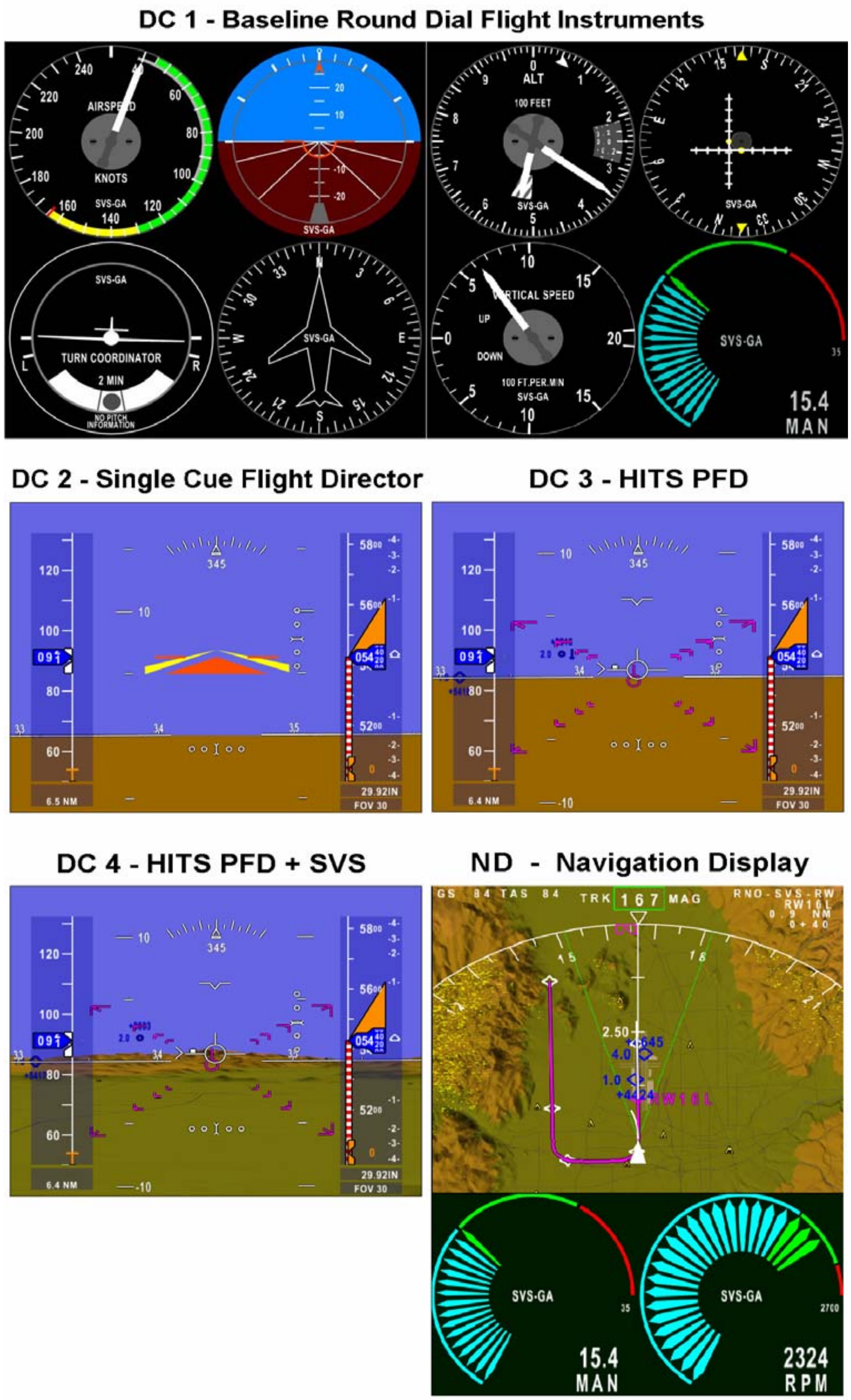

Figure 1. SVS-ES Display Concepts

Figure 2. The NASA Integrated Flight Deck (IFD) Configured For SVS-ES Experiment

2.4 Scenarios

The evaluation pilots were required to fly three different styles of approaches at RNO; two conventional that reflected actual operations at RNO, and one advanced reflecting potential future operations. The conventional maneuvers included a Visual Flight Rules (VFR) box pattern approach to runway 16L at RNO conducted in VMC out-the-window visibility and an Instrument Landing System (ILS) approach to 16R conducted in IMC out-the-window visibility. A wind model, that based wind direction and magnitude on the aircraft's altitude, as well as light to moderate turbulence, was incorporated for this experiment in an attempt to model moderately challenging atmospheric conditions. Every EP, regardless of experience or rating, flew both the VFR and ILS approaches once with each of the four experimental display concepts. The advanced maneuver utilized the same flight path flown for the conventional VFR box pattern approaches; however, this "VMC-like" approach was flown in IMC out-the-window visibility. Again, every EP, regardless of experience or rating flew the VMC-like approach with each of the experimental display concepts except the BRD. The BRD DC runs were excluded from testing with the VMC-like approach due to its inability to provide the pilot with the enough information to safely and accurately navigate the box pattern in IMC thereby making them unsuitable for such a maneuver. The VMC-like approach represents a potential future operation that could one day be enabled by advanced cockpit displays, such as SVS. A total of thirty-three scenarios (combination of maneuver, display concept, and pilot category) were investigated as part of this research. Reno-Tahoe International Airport (RNO), in Reno, Nevada, was chosen as the simulation environment for this research.

All scenarios were initialized on the desired flight path at 90 kts airspeed with the flaps retracted. When flying the graded segments for the ILS and VMC-like maneuvers, the EP's were instructed to use 20 degrees of flaps, and to minimize lateral and vertical path errors and decelerate to and maintain 80 kts of airspeed to the best of their ability. For the VFR maneuver, the EP's were instructed to fly visually while maintaining 90 kts of airspeed from scenario initialization until they were abeam the runway numbers; they were then to select 20 degrees of flaps and decelerate to and maintain 80 kts of airspeed for the remainder of the scenario, in order to be consistent with traditional VFR box pattern operations. A god's-eye-view of the two different approach paths utilized for the three maneuvers are provided below (Figure 3).

Each evaluation pilot flew a total of 11 data collection runs in the IFD during their one day participation in the experiment. During the experiment the order of maneuver and display type was randomized across EP's to minimize any learning or fatigue effects in the data. In addition, the EP's were required to monitor simulated Air Traffic Control (ATC) during the approaches and had to be prepared to provide a response to a workload-inducing memory recall question at the end of each data run about the traffic situation during the approach. For the ILS and VMC-like maneuvers the EP's were also instructed to be prepared to declare, and initially execute, a missed-approach if they were unable to see the airport environment in the out-the-window scene when they arrived at the 200 ft Above Ground Level (AGL) decision height (DH) chosen for this experiment.

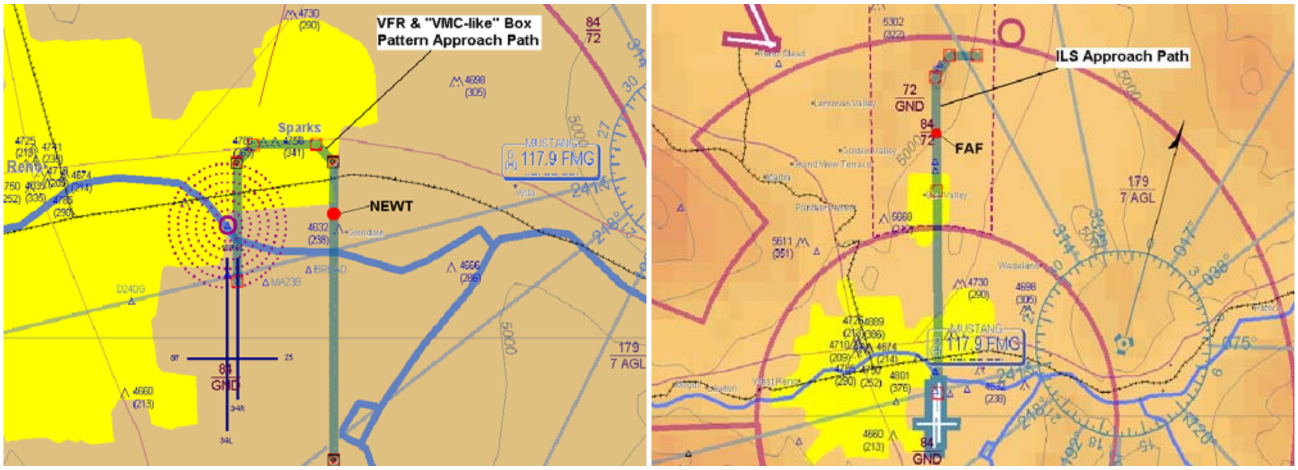

Figure 3. SVS-ES Approach Paths

2.5 Training

A substantial amount of time was allocated to brief the EP's and familiarize them with the IFD, experimental display concepts, and maneuvers. General "air work" with a Certified Flight Instructor Instrument-Rated (CFII) was flown with all four of the display concepts in order to familiarize the subjects with the general display layouts as well as the control feel provided by the hydraulic control loader system of the IFD. After the "air work" was completed, practice VFR and ILS approaches were flown with the same DC's and CFII to familiarize the EP's with the basic procedures for the approach maneuvers they would fly during data collection. Overall, approximately 2 hours of classroom style briefings and 2 hours of simulator dual instruction were performed for each EP for this experiment.

2.6 Performance Criteria

The SVS-ES experiment was designed to provide relative safety related data about current and potential future operating conditions. Currently, a common method to determine acceptably safe operating conditions is the FAA Practical Test Standards (PTS). In order to earn an instrument rating from the FAA, a pilot must demonstrate the ability to maintain flight path and airspeed errors within the acceptable limits outlined in the PTS. As a result, the accepted safe condition, herein referred to as "safety tunnel", for the ILS approach maneuver was based on the FAA PTS criteria of ¾ scale maximum deflection (1.5 dots) of localizer and glide slope error and +/- 10 kts of airspeed error (target airspeed was 80 kts) required to pass an IFR flight test. In addition, to facilitate these initial data analyses and provide a uniform evaluation criterion for the IMC maneuvers, the modified form of the ILS safety tunnel, including airspeed error, was applied to the VMC-like maneuver. Using linear flight path error and range along flight path to touchdown, ILS-like range-varying angular error was calculated for the VMC-like maneuver. For the VFR maneuver, the accepted safe condition was created based on an assumed amount of lateral and vertical error, acceptable to an instructor pilot, and was based on the VFR PTS. The resulting VFR approach "safety tunnel" had dimensions of +/- 0.5 nm laterally and +/- 250 ft vertically. The VFR approach safety tunnel gradually tapered from its full size during the final approach segment, beginning at approximately 1.5 nm from the touchdown zone, narrowing from its nominal size down to 150 ft laterally and 0 ft vertically at the runway 16L touchdown location. Airspeed error was not applied to the VFR safety tunnel for these initial analyses. EP's were instructed that their technical performance would be determined as the percentage of time during the approach they spent within the invisible safety tunnel for all approach maneuvers. Finally, the VFR safety tunnel begun at the scenario initialization point, the ILS safety tunnel began at FAF waypoint, and the VMC-like safety tunnel began at the NEWT waypoint (Figure 3).

3. DATA

3.1 Subjective Data

After each data collection run, the EP's were asked to complete a battery of subjective run questionnaires using a tablet PC. The run questionnaire battery consisted of traditional cockpit display research questionnaires: the 3-dimensional

Situation Awareness Rating Technique (3-D SART) (Taylor & Selcon, 1991) which measures SA; the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988) which measures workload, and Haworth-Newman Avionics Display Readability Modified Cooper Harper Scale (MCH) (Chiappetti, 1994; Gawron, 2000) which measures information display readability. At the end of the experiment the EP's completed several paired comparisons and general questions, again using the tablet PC. This exit questionnaire paired comparison task was used to gather the EPs' opinions on the relative contribution of situation awareness, workload, and information display readability to the safety associated with flying an approach. The paired comparison data was used to calculate subjective component weights for the individual post-run questionnaires that were combined into the prototypes of an Equivalent Safety Metric (ESM) evaluated in this research. In addition, all of the questionnaire data was collected using 0-100 pt continuous scales.

3.2 Objective Data

During each data collection run, a wide variety of flight technical error (FTE) data were recorded in order to quantify the EP's performance while flying the approach maneuvers. However, for this initial metric development effort, only the percentage of time within the safety tunnel for each approach maneuver was included in the analysis and is referred to as "% Time Safety Tunnel". In order to facilitate these initial analyses, the portion of the ILS approach maneuver from the Final Approach Fix (FAF) to the Decision Height (DH), where the EPs were controlling airspeed to 80kts, was used to calculate the EP's "% Time Safety Tunnel" score. Similarly, for the VMC-like maneuver the portion of the approach from the NEWT waypoint at the end of the downwind segment to the DH was used to calculate the EP's "% Time Safety Tunnel" score. The location of the FAF and NEWT waypoints can be seen in Figure 3. For the VFR maneuver the entire approach was used to calculate the EP's "% Time Safety Tunnel" score.

Metric	Components	Weights (%)	Scale
ESM-1	Workload	100	0 - 100
ESM-2	Workload	31	0 - 100
	SA	69	
ESM-3	Workload	33	0 - 100
	SA	37	
	Readability	30	
ESM-4	Workload	33	0 - 100
	SA	37	
	Readability	30	
	% Time Safety Tunnel	100	

Table 2. Component Weights For ESM Prototypes

3.3 Preliminary Equivalent Safety Metrics

Four preliminary Equivalent Safety Metric's (ESM) were composed and evaluated as part of this research. The first three preliminary metrics were purely subjective and built solely from combinations of the three subjective questionnaires that composed the post-run subjective questionnaire battery. Corresponding weights were calculated from the paired comparison task performed by the EP's in the exit questionnaire using a non-parametric analysis. The first preliminary ESM (ESM-1) was composed solely of subjective workload ratings gathered from the EP's using the NASA-TLX post run subjective questionnaire. Starting with only subjective workload reflects a basic approach towards assessing relative safety. The second preliminary ESM (ESM-2) was composed of the weighted sum of the 3-D SART and NASA-TLX post-run subjective questionnaire data. Adding the SA component enables quality of the pilot's mental model to be incorporated into the safety assessment. It is important to note that for constructing the preliminary metrics, both the NASA-TLX and MCH raw data was "directionalized" so that favorable responses would be indicated by high values. Proceeding in this manner enables the analysis of the combined preliminary metrics to be more intuitive and easily interpreted. Furthermore, it has been theorized in this research that the information display readability of advanced cockpit displays implies a cost or workload associated with incorporating critical flight information into the pilot's mental model of the situation (SA). Therefore, using the "directionalized" weighted data, the third preliminary ESM can be simply be defined as: ESM-3 = Readability (R) + Awareness (A) + Workload (W). The fourth and final prototype ESM (ESM-4) combines ESM-3 with the "% Time Safety Tunnel" scores for all scenarios. Including actual pilot performance adds a meaningful quantitative dimension to the metric. In addition, ESM-4 has been normalized to a 100 pt scale to make it easier to evaluate along side the other three purely subjective prototype ESMs. The table above (Table 2) details the components of the four prototype ESMs and their corresponding weights and scale ranges.

4. RESULTS

The subjective data analyses utilized repeated measures Analysis of Variance (ANOVA) and Student-Newman-Keuls' post-hoc tests to analyze the three individual questionnaires that compose the subjective post-run questionnaire battery. Repeated measures ANOVA's and Student-Newman-Kuels' post-hoc tests were also used to analyze the "% Time Safety Tunnel" and the four prototype ESM's Finally, the evaluation of the four ESM prototypes relied on a comparison of critical means representative of known 'safe' and 'unsafe' scenarios, described later in this research.

4.1 NASA Task Load Index Results

Considering the data from all four DC's flown with the conventional maneuvers (i.e. VFR and ILS maneuvers), a significant effect for Maneuver Type was found to exist with respect to TLX scores, $F(1,22) = 77.19$. The mean TLX workload score for the VFR (40.35) maneuver was significantly lower than the mean TLX score for the ILS (52.37) maneuver. Excluding the BRD DC data and looking only at the three advanced display concepts flown with all three maneuvers, again the VFR (40.19) approach had a significantly lower workload score than the VMC-like (46.76) and ILS approach (48.84).

There were no significant effects found for Pilot Category with respect to TLX scores for any of the maneuvers ($p > .05$). However, the IFR pilots (35.47) reported substantially lower workload than the H-IFR pilots (45.19) and VFR pilots (40.38) for the VFR maneuver.

A significant effect for Display Concept was found to exist with respect to TLX scores for the conventional maneuvers (i.e. VFR and ILS maneuvers), $F(3,72) = 9.03$. The Baseline Round Dial DC (52.02) was found to be significantly worse than all three other more advanced DC's in terms of workload. There were no significant differences in workload detected between the three advanced DC's during the conventional maneuvers. Excluding the BRD DC data and looking only at the three advanced display concepts flown with all three maneuvers, there were again no significant differences in workload.

For the conventional approach maneuvers (i.e. VFR and ILS maneuvers) flown with all four experimental DC's, the interaction of Display Concept and Maneuver Type with respect to TLX scores was found to be significant, $F(3,66) = 14.17$. The ILS maneuver flown with the BRD (63.08) was found to have a significantly higher workload score than the ILS maneuvers flown with the SCFD (49.80), HITS PFD (49.04), or SVS + HITS PFD (48.48), as well as any of the four DC's flown in the VFR maneuver. When considering only the three advanced DC's across all three maneuver types, the interaction of Display Concept and Maneuver Type was not found to be significant with respect to TLX scores($p > .05$). Thus, the workload for the three advanced display concepts was not found to be significantly different from one another regardless of OTW visibility.

There are several test conditions of interest within the Display Concept and Maneuver Type interaction effect. The ILS maneuver flown with the BRD (63.08) was found to have significantly higher workload ratings than all other test conditions. Interestingly, the VMC-like maneuver flown with SVS + HITS (46.08) was not found to have significantly a higher workload rating than the VFR maneuver flown with BRD (40.96).

4.2 Situation Awareness Rating Technique Results

Considering the data from all four DC's flown with the conventional maneuvers, a significant effect for Maneuver Type was found to exist with respect to 3-D SART scores, $F(1,22) = 76.75$. The mean 3-D SART score for the VFR (100.35) maneuver (with clear out-the-window visibility) was significantly higher (better) than the score for the ILS (51.37) maneuver. Excluding the BRD DC data and looking only at the three advanced display concepts flown with all three maneuvers, a significant effect for Maneuver Type was again detected with respect to 3-D SART scores, $F(1,22) = 27.85$. Again, the VFR (105.57) approach had significantly higher mean 3-D SART scores than the VMC-like (64.00) and ILS (66.72) maneuvers.

Pilot Category was not found to be significant for any of the maneuvers with respect to 3-D SART scores ($p > .05$).

When considering the data for only the conventional maneuvers, there was a significant main effect found for display concept (DC) with respect to situation awareness, $F(3,72) = 12.29$. The Baseline Round Dial DC (47.54) was rated significantly lower in terms of SA than the three other more advanced display concepts for the VFR and ILS approaches. Excluding the BRD DC data and looking only at the three advanced display concepts flown with all three maneuvers, the SVS + HITS DC (85.79) was rated significantly higher in terms of SA than the other display concepts. The SCFD (74.03) and HITS PFD (74.96) were not found to be significantly different from one another in terms of SA for the three maneuvers.

For the conventional approach maneuvers (i.e. VFR and ILS maneuvers) flown with all four experimental DC's, the interaction of Display Concept and Maneuver Type with respect to 3-D SART scores was found to be significant, $F(3,66) = 7.12$. The ILS maneuver flown with the BRD (6.00) was found to have a significantly lower mean 3-D SART score than ILS maneuvers flown with the SCFD (58.88), HITS PFD (65.76), or SVS + HITS PFD (75.04), as well as any of the four DC's flown in the VFR maneuvers. In addition, the mean 3-D SART score for the ILS maneuver flown with the SVS + HITS PFD (75.04) DC was not found to be significantly different from the VFR maneuver flown with the BRD (84.80) DC.

When considering only the three advanced DC's flown for all three maneuver types, the interaction of Display Concept and Maneuver Type was found to be significant with respect to 3-D SART scores, $F(4,88) = 2.72$. The mean 3-D SART scores for the VFR maneuver flown with any of the advanced DC's were higher than the mean 3-D SART scores for either the VMC-like or ILS maneuvers flown with any of the advanced DC's.

Of the several test conditions of interest within the Display Concept and Maneuver Type interaction effect, the score for the ILS maneuver flown with the BRD (6.0) was found to be significantly lower than all other test condition scores with respect to SA. Conversely, the score for the VFR maneuver flown with the BRD (84.80) was not found to be significantly different from any scores of the SVS display concepts flown in the ILS or VMC-like maneuvers with respect to SA.

4.3 Haworth-Newman Avionics Display Readability Scale Results

Considering data from all four DC's flown with the conventional maneuvers, a significant effect for Maneuver Type was found to exist with respect to MCH scores, $F(1,22) = 8.59$. The mean MCH score for the ILS (34.85) maneuver was significantly higher (worse) than the mean MCH score for the VFR (30.42) maneuver. Excluding the BRD DC data and looking only at the three advanced display concepts flown with all three maneuvers, a significant effect for Maneuver Type was not detected with respect to MCH scores ($p > .05$).

Pilot Category was not found to be significant for any of the maneuvers with respect to MCH scores ($p > .05$).

When considering the data for only the conventional maneuvers, there was a significant effect for display concept found to exist with respect to MCH scores, $F(3,72) = 11.13$. The Baseline Round Dial DC (43.20) MCH score was found to be significantly worse than all three other more advanced DC's for the VFR and ILS maneuvers. Excluding the BRD DC data, the SVS + HITS DC (26.53) was rated significantly better (i.e lower) in terms of readability than the other three display concepts. The SCFD (31.33) and HITS PFD (30.13) were not found to be significantly different from one another in terms of readability.

For the conventional approach maneuvers (i.e. VFR and ILS maneuvers) flown with all four experimental DC's, the interaction of Display Concept and Maneuver Type with respect to MCH scores was found to be significant, $F(3,66) = 8.40$. The ILS maneuver flown with the BRD (50.40) was found to have significantly higher (worse) mean MCH score than the ILS maneuvers flown with the SCFD (31.20), HITS PFD (30.40), or SVS + HITS PFD (26.80), as well as any of the four DC's flown in the VFR maneuver. However, when considering only the three advanced DC's across all three maneuver types, the interaction of Display Concept and Maneuver Type was not found to be significant with respect to MCH scores ($p > .05$). Logically, OTW visibility does not have an impact on the display readability scores for the three advanced DC's.

Of the several test conditions of interest within the Display Concept and Maneuver Type interaction effect, the VFR (30.60) and ILS (50.40) maneuvers flown with the BRD were found to have significantly worse readability than all other test conditions

4.4 Percent Time Within Safety Tunnel Results

Considering data from all four DC's flown with the conventional maneuvers, a significant effect for Maneuver Type was found to exist with respect to % Time Safety Tunnel scores, $F(1,22) = 41.39$. The mean % Time Safety Tunnel score for the VFR (98.23) maneuver was significantly higher than the mean % Time Safety Tunnel score for the ILS (85.82) maneuver. Excluding the BRD DC data and looking only at the three advanced display concepts flown with all three maneuvers, a significant effect for Maneuver Type was again detected with respect to % Time Safety Tunnel scores, $F(1,22) = 19.81$. Again, the VFR (98.40) approach had significantly higher mean % Time Safety Tunnel scores than the VMC-like (82.04) and ILS (89.93) maneuvers.

Pilot Category was found to be significant for the ILS maneuver with respect to % Time Safety Tunnel scores, $F(2,22) = 11.77$. As one might expect, the VFR pilot group (70.16) was found to have a significantly lower mean % Time Safety Tunnel score for the ILS maneuver then both the IFR (91.84) and H-IFR (95.47) pilot groups. There were no effects detected for Pilot Category for the VFR or VMC-like maneuvers with respect to % Time Safety Tunnel scores ($p > .05$).

Considering data from all four DC's flown with the conventional maneuvers, no significant effect for Display Concept (DC) was found to exist with respect to % Time Safety Tunnel scores, nor was a significant effect found when excluding the BRD DC data and looking only at the three advanced display concepts flown with all three maneuvers.

For the conventional approach maneuvers (i.e. VFR and ILS maneuvers) flown with all four experimental DC's, the interaction of Display Concept and Maneuver Type with respect to % Time Safety Tunnel score was found to be significant, $F(3,66) = 3.99$. The ILS maneuver flown with the BRD (74.32) was found to have a significantly lower mean % Time Safety Tunnel scores than ILS maneuvers flown with the SCFD (89.88), HITS PFD (92.52), or SVS + HITS PFD (88.12), as well as any of the four DC's flown in the VFR maneuver. However, when considering only the three advanced DC's across all three maneuver types, the interaction of Display Concept and Maneuver Type was not found to be significant with respect to % Time Safety Tunnel scores ($p > .05$).

Of the several test conditions of interest within the Display Concept and Maneuver Type interaction effect, the ILS maneuver flown with the BRD (74.32) was found to have a significantly lower mean % Time Safety Tunnel score than all other test conditions. The % Time Safety Tunnel score for the VFR maneuver flown with BRD was somewhat higher than the scores for the VMC-like maneuvers with any of the advanced display concepts.

5. DISCUSSION

5.1 Evaluation Of the Four Preliminary Equivalent Safety Metrics Based on Critical Deltas

In order to evaluate the prototype ESM's it has been theorized that an effective ESM would be capable of revealing meaningful differences between known 'safe' and 'unsafe' scenarios - scenarios that can be associated with documented accident rates. According to the current FAA regulations, VFR rated pilots are logically prohibited from flying in IMC, which includes executing an ILS approach in IMC. Also, the occurrence of non-instrument rated pilots operating in IMC is a leading causal factor in the high rate of fatal GA accidents. Therefore, in the SVS-ES experiment, the situation of VFR rated pilots flying an ILS approach with the BRD DC was categorized as an 'unsafe' scenario. Similarly, IFR rated pilots, according to the FAA regulations, are legally qualified to fly an ILS approach in IMC with the BRD DC, categorizing this scenario as a 'safe' scenario, though one that still possesses an elevated rate of accidents. Furthermore, given their experience and resulting low-rate of accidents per flight hour, it was assumed that high-time IFR rated pilots (H-IFR Pilot Category) flying the VFR approach with the BRD DC constitutes a second, potentially 'most safe' scenario (categorized herein as 'safer'). Therefore, a comparison of the means for these 'safe' and 'unsafe' scenarios, with documented and substantially different real-world accident rates, constituted a reasonable method for evaluating the prototype ESM's constructed as part of this research. The table below (Table 3) details the comparisons of the known 'safe' and 'unsafe' scenarios tested in the SVS-ES experiment and includes the calculated percent change of the measured means with respect to the three aforementioned scenarios.

Based on the comparison of known 'safe' and 'unsafe' scenarios performed with the four preliminary ESM's it is apparent that ESM-1, composed solely of subjective NASA-TLX workload estimates, does not constitute a suitable method of measuring the safety associated with flying an approach compared to the other three preliminary ESM's due to its weak sensitivity. ESM-2, the weighted combination of subjective NASA-TLX workload and 3-D SART SA estimates, shows a substantial improvement over ESM-1 in terms of sensitivity. However, ESM-3, the weighted combination of subjective NASA-TLX workload, 3-D SART SA, and Haworth-Newman MCH readability estimates (R+A+W) appears to be slightly more capable of maximizing the differences between the known 'safe' and 'unsafe' scenarios than ESM-2. Finally, adding the objective "% Time Safety Tunnel" data to the subjective ESM-3 data seems to show the largest sensitivity, especially for describing the critical comparison between the 'unsafe' scenario of VFR rated pilots flying an ILS approach with the BRD and the 'safe' scenario of IFR rated pilots flying the ILS approach with the BRD. Therefore, it is believed that in order to adequately compare the safety associated with flying an approach with round dial flight instruments to other more advanced flight information displays, including a combination of subjective and objective data measurements appears reasonable. Stated another way, inclusion of FTE data as described herein, adds to the sensitivity of the ESM-4, the prototype metric.

		Scenario	ESM-1	ESM-2	ESM-3	ESM-4
Means	1 'unsafe'	VFR Pilots Flying ILS Approach with BRD	37.13	15.01	21.06	34.66
Means	2 'safe'	IFR Pilots Flying ILS Approach with BRD	40.38	21.32	30.54	54.33
Means	3 'safer'	H-IFR Pilots Flying VFR Approach with BRD	54.89	41.20	55.65	76.38
Comparison of Means		Critical Δ1 (Mean 2 – Mean 1) 'safe'-'unsafe'	3.25	6.31	9.48	19.67
Comparison of Means		Critical Δ1 (%) 100% x ('safe'-'unsafe')/'safer'	5.92	15.31	17.04	25.75
Comparison of Means		Critical Δ2 (Mean 3 – Mean 1) 'safer'-'unsafe'	17.76	26.19	34.59	41.72
Comparison of Means		Critical Δ2 (%) 100% x ('safer'-'unsafe')/'safer'	32.36	63.57	62.15	54.62
		ESM Sensitivity Critical Δ1 (%) + Critical Δ2 (%)	38.28	78.87	79.20	80.39

Table 3. Comparison of Known 'Safe' & 'Unsafe' Scenarios

5.2 Discussion Of the Prototype Equivalent Safety Metric

After having gained confidence in the ESM-4 prototype metrics' ability to adequately describe the difference between the known 'safe' and 'unsafe' scenarios, it was applied to the remaining scenarios tested during the SVS-ES experiment. A simple ANOVA was performed on the ESM-4 scores calculated for each of the data collection runs using the variable "Scenario" - a combination of Display Concept, Maneuver, and Pilot Category.

As mentioned previously, merging the % Time Safety Tunnel FTE data with ESM-3 into the final prototype ESM (ESM-4) did provide some enhanced ability to differentiate between the VFR and IFR pilots performing the ILS approach with the BRD (the 'unsafe' and 'safe' scenarios), which was an expected and desired result. However, ESM-4 then became insensitive to other results which had been reflected in ESM-3, such as the effect of SVS terrain. The reason for this change in sensitivity is clear, since the FTE results indicated that there were no significant differences between the SCFD, HITS PFD, and SVS + HITS PFD, regardless of pilot group or maneuver. Combining % Time Safety Tunnel FTE data with the subjective data for EMS-4 overwhelmed the contribution of the individual components of the subjective half of the metric. Because 50% of ESM-4 was % Time Safety Tunnel, it was roughly having three times the individual influence of SART, TLX, and MCH on the overall ESM-4 scores calculated for the scenarios. Therefore, the rather large influence of FTE data on the EMS-4 scores disabled the metric's ability to differentiate between the advanced display concepts; the SCFD, HITS PFD, and SVS + HITS PFD. The three advanced display concepts were all statistically similar based on the mean % Time safety Tunnel scores for all three maneuvers but were significantly different from one another subjectively for the VMC-like and ILS maneuvers. Further work regarding the appropriate influence of FTE data on the resulting metric is warranted and will be performed during the future planned development of this "Equivalent Safety Metric".

A significant effect for "Scenario" was found to exist as indicated by ESM-4, $F(30,220) = 8.35$. The ILS maneuver flown by the VFR pilot group with the BRD DC scenario was found to have the lowest mean ESM-4 score and was significantly different from all of the other scenarios tested during the SVS-ES experiment. Also, the ILS maneuver flown by the IFR pilot group with the BRD DC scenario was found to have the second lowest mean ESM-4 score and was not found to be significantly different from either the VMC-like or ILS maneuvers flown by the VFR pilot group with the SVS+HITS DC. This result indicates that VFR pilots can fly ILS and VMC-like maneuvers with SVS displays to a level of safety comparable to today's ILS approach in IMC flown with round dial flight instruments by IFR-rated pilots. Furthermore, it is worth noting that 4 of the 8 ILS maneuver with BRD data collection runs flown by the VFR pilot groups ended prematurely. Three of the four incomplete data collection runs were stopped because the EP's got disoriented and were unable to find the airport using the BRD DC (localizer and glide slope needles) and the fourth incomplete data collection run ended in a Controlled Flight Into Terrain (CFIT) accident. Conversely, all 8 VFR pilots were able to successfully fly every ILS and VMC-like maneuver to conclusion (i.e. landing or declared missed approach) using all three of the other more advanced display concepts tested in the SVS-ES experiment. Another interesting result found from the ANOVA performed on the ESM-4 scores by "Scenario" is that means for both the IFR and H-IFR pilot groups flying the VFR maneuver with the BRD DC were not found to be significantly different then the IFR and H-IFR pilot groups flying the either the ILS or VMC-like maneuvers with the SVS+HITS DC. However, this result is not true for the SCFD or HITS PFD advanced DC's. The table and bar chart above (Figure 4) provides a graphical representation of the ESM-4 means for all 33 scenarios tested in the SVS-ES experiment and includes the results of the Student-Newman-Kuels' post-hoc test.

A significant effect for Display Concept was found to exist with respect to ESM-4 scores for the VFR maneuver, $F(3,66) = 3.13$. The baseline round dials (76.05) was found to have a significantly lower ESM-4 score than the HITS PFD (80.56) and SVS + HITS PFD (79.72) display concepts. Pilot Category did not have a significant effect on ESM-4 scores for the VFR maneuver ($p > .05$). A significant effect for Display Concept was also found to exist with respect to ESM-4 for the ILS maneuver, $F(3,66) = 17.79$. The baseline round dials (50.78) was found to have a significantly lower ESM-4 score for the ILS maneuver than all three more advanced display concepts; SCFD (68.92), HITS PFD (71.14), and SVS + HITS PFD (70.41). In addition, Pilot Category was found to have a significant effect on ESM-4 scores for the ILS maneuver, $F(2,22) = 10.35$. The VFR (56.24) pilot group was found to have a significantly lower ESM-4 score for the ILS maneuver than both the IFR (69.88) and H-IFR (69.845) pilot groups. The interaction of Display Concept and Pilot Category was not found to be significant for the ILS maneuver with respect to ESM-4 scores ($p > .05$). Finally, there were no significant effects detected for Display Concept or Pilot Category with respect to ESM-4 scores for the

VMC-like maneuver (p > .05), indicating that the advanced displays enabled VFR pilots to perform similarly to the other two pilot groups with the three advanced DC's investigated for the VMC-like maneuver.

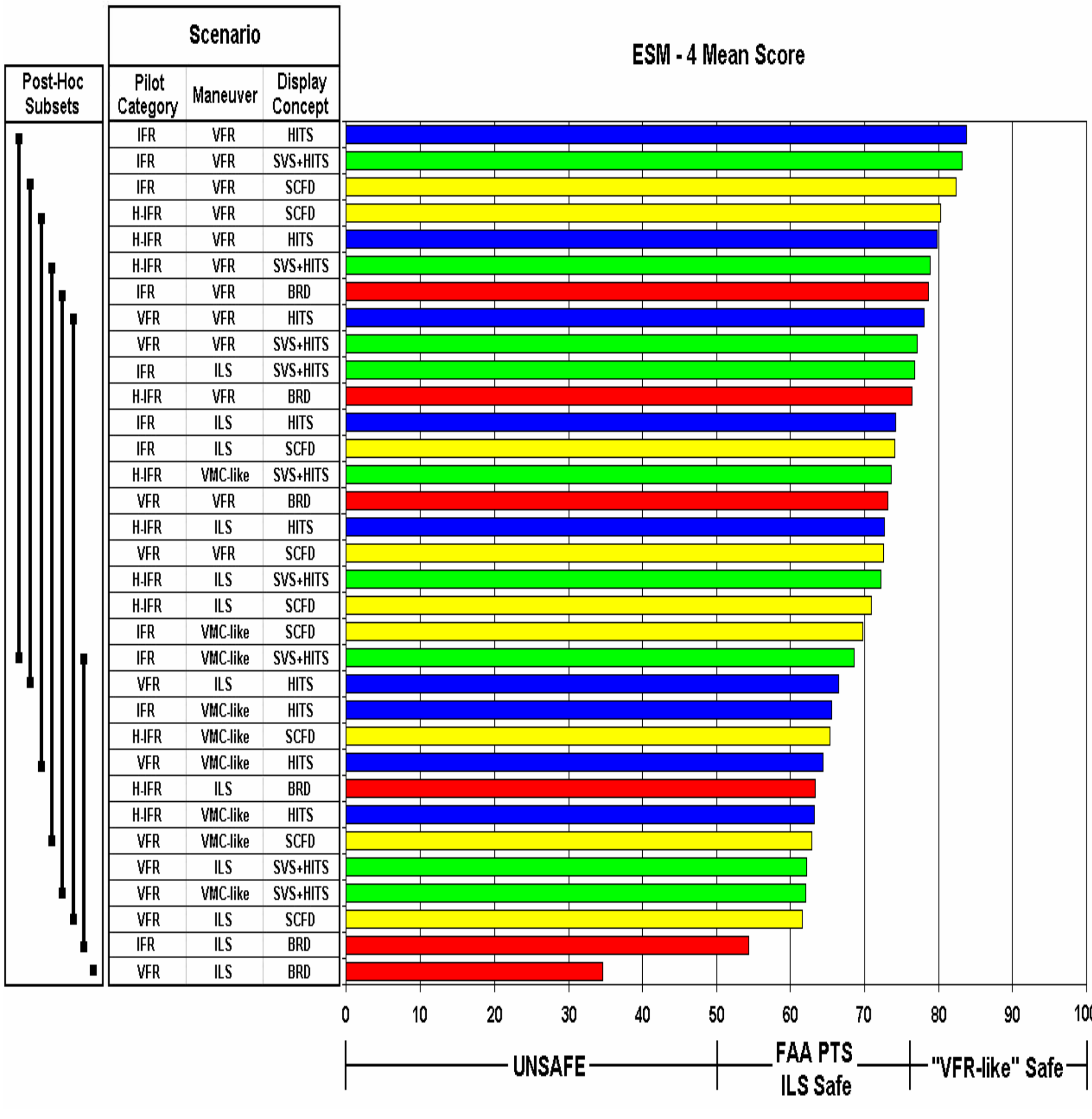

Figure 4. ESM-4 Mean Scores by Scenario with S-N-K Post-Hoc Subsets

6. CONCLUSIONS

Previous SVS research performed by NASA (Glaab & Hughes, 2003; Kramer et al, 2003; Takallu et al., 2004; Wong et al., 2004; Prinzel et al, 2004), academia, and industry, have all indicated that a pilot's situation awareness can be substantially improved when flying in IMC by providing a SVS terrain background on a primary flight display. Furthermore, by overlaying the digital representation of real-world terrain with an intuitive pathway based guidance symbology, greater precision is achievable, regardless of a pilot's level of experience, without having a negative impact on the pilot's workload. The results from the SVS-Equivalent Safety simulation experiment have confirmed the results found by previous SVS research and extend that research in a more comprehensive manner. For the first time, several candidate primary flight display and navigation display packages, including a fully integrated SVS package, have been comprehensively compared to the conventional round dial GA flight instruments in VMC and IMC. Furthermore, the concept of equivalent safety has been applied towards establishing the safety and operational capabilities of SVS displays. As a result, the continued development of an "Equivalent Safety Metric" has been identified as an integral part of the ongoing NASA Langley SVS-GA simulation and flight test research program.

The prototype ESM discussed herein combined weighted subjective pilot evaluations of display readability, situation awareness, and workload, with an objective measurement of performance based on the accepted safe standard for flight in IMC, the FAA Practical Test Standards (PTS), into a means of determining the relative safety of a candidate display concept through a comparison to the current operationally accepted standard for GA. When applied to the data collected during the SVS-ES simulation experiment, the prototype ESM appears to be capable of distinguishing between the known 'unsafe' scenario of VFR pilots flying an ILS approach in IMC with the BRD. Furthermore, by being progressive enough to consider the modifications to today's GA IMC operations that SVS could possibly one day enable, the inclusion of the "VMC-like" maneuver in the SVS-ES experiment provided the researchers with the opportunity to investigate the potential operational benefits of SVS displays as well. The results of the statistical analysis performed on the prototype ESM scores by "scenario" indicate there are no statistically significant differences between the means calculated for IFR or H-IFR pilots flying either the ILS or VMC-like maneuvers in IMC with SVS and any pilot group flying the VFR maneuver with the round dial flight instruments. This statement is not true for the other two advanced display concepts (i.e. the SCFD and PFD with HITS). Therefore, although it is important to remember that this paper constitutes an initial metric development effort that will be augmented with flight test data, within the context of this experiment and it's metrics, SVS displays made IMC operations appear equivalent to those operations conducted in VMC.

The advent of advanced affordable electronic cockpit displays for general aviation (GA) aircraft is rapidly approaching. Major GA airframe manufacturers are teaming with the leading avionics producers to revolutionize the pilot interface for the 21st century. However, currently there is no clearly defined certification process or method of evaluation to determine whether an "equivalent level of safety" will be maintained with advanced information displays such as synthetic vision. The prototype metric discussed in this research appears to represent a plausible initial method of describing the relative safety associated with the candidate display concepts tested in the SVS-ES simulation experiment. However, the relative influence of FTE data warrants further development. The prototype ESM constructed and evaluated in this research represents the first step toward the confirmation and validation of a robust technique and has demonstrated an initial ability to evaluate a variety information display types in terms of their relative contribution to safety and operational flexibility while flying both conventional and potential future approach maneuvers. The continued development of the prototype "Equivalent Safety Metric" has been identified as an integral part of the ongoing NASA Langley SVS-ES simulation and flight test research

The next phase of the SVS-ES experiment is to conduct a flight test utilizing the NASA LaRC C-206 research aircraft to evaluate the efficacy of the advanced displays in the real world comparison to "round dials". Results from this flight test will aid the interpretation and application of results of this research, confirm the test techniques employed for the SVS-ES simulation, as well as provide additional data for the further development of the prototype ESM. Overall, the SVS-ES experiment provides information regarding the relative safety of SVS, and other advanced displays, assuming normal operating conditions. Additional future research needs to be conducted to consider various system failure modes of the advanced display concepts to effectively capture the safety associated with the various display concept and maneuver combinations.

REFERENCES

Bartolone, A., Hughes, M., Takallu, M., and Wong, D., *Experiment Results for Symbology Development for General Aviation Synthetic Vision Primary Flight Displays for the Approach and Missed Approach Modes of Flight*, Proceedings 48th Annual Meeting of the Human Factors and Ergonomics Society, 2004

Chiappetti, C., *Evaluation of the Haworth-Newman Avionics Display Readability Scale*, Naval Postgraduate School Thesis, Monterey, CA, 1994

Endsley, M., *Situation Awareness in Aviation Systems*, In D.J. Garland, J.A.Wise, & V.D. Hopkin (Eds.) Handbook of Aviation Human Factors, Lawrence-Erlbaum Assoc., Mahwah, NJ, 1999

Gawron, V., *Human Performance Measures Handbook*, Lawrence Erlbaum Assoc., Mahwah, NJ, 2000

General Aviation Manufacturers Association (GAMA), *GAMA Publication #12: Recommended Practices and Guidelines for an Integrated Cockpit/Flight Deck in a 14 CFR Part 23 Certificated Airplane,* GAMA, Washington, D.C., 2004

Glaab, L., and Hughes, M., *Terrain Portrayal for Head-Down Displays Flight Test*. 22nd Annual Digital Avionics Systems Conference (DASC). Indianapolis, IN, 2003

Glaab, L., and Hughes, M., *Terrain Portrayal for Head-Down Displays Simulation Experiment Results*. 22nd Annual Digital Avionics Systems Conference (DASC). Indianapolis, IN, 2003

Greer, D., Shivers, J., Fink, M., and Carter, R., *Wind-Tunnel Investigation of Static Longitudinal and Lateral Characteristics of a Full-Scale Mockup of a Light Single-Engine High-Wing Airplane*, NASA TN D-7149, 1973

Hart, S., and Staveland, L., *Development of the NASA Task Load Index (TLX): Results of Empirical and Theoretical Research*, In P.A. Hancock & N. Meshkati (Eds.), Human Mental Workload, Elsevier Publishing, Amsterdam, Holland, 1988

Kramer, L., Prinzel, L., Bailey, R., & Arthur, J., *Synthetic Vision Enhances Situation Awareness and RNP Capabilities for Terrain-Challenged Approaches*, Proceedings of the American Institute of Aeronautics and Astronautics Third Aviation Technology, Integration, and Operations Technical Forum, AIAA 2003-6814, 2003

NASA Langley Research Center (LaRC): Staff of the Flight Dynamics Branch, *Measurement of the Handling Characteristics of Two Light Airplanes*, NASA TP-1636, 1980

Prinzel, L., Arthur, J., Kramer, L., and Bailey, R., *Pathway Concepts Experiment for Head-Down Synthetic Vision Displays*, Proceedings of SPIE, Enhanced and Synthetic Vision Conference, 2004

Prinzel, L, Comstock, J., Glaab, L., Kramer, L., Arthur, J., & Barry, J., *The Efficacy of Head-Down and Head-Up Synthetic Vision Display Concepts for Retro- and Forward-Fit of Commercial Aircraft,* International Journal of Aviation Psychology, 17, 2004

Riley, D., Brandon, J., and Glaab, L., *Piloted Simulation Study of an ILS Approach of a Twin-Pusher Business Commuter Turboprop Aircraft Configuration*, NASA TM-4516, 1994

Roskam, J., *Airplane Flight Dynamics and Automatic Controls Part-1*, Roskam Aviation and Engineering Corp.,1979

Takallu, M., Wong, D., Bartolone, A., Hughes, M., and Glaab, L., *Interaction Between Various Terrain Portrayals and Guidance/Tunnel Symbology Concepts for General Aviation Synthetic Vision Displays During a Low En-Route Scenario*, Proceedings of the 23rd Digital Avionics Systems Conference, 2004

Taylor, R., and Selcon, S., Subjective Measurement of Situation Awareness, Proceedings of the 11th Congress of the International Ergonomics Association, 1991

Wong, D., Takallu, M., Hughes, M., Bartolone, A., and Glaab, L., *Simulation Experiment for Developing the Symbology for the Approach and Missed Approach Phases of Flight for Head-Down Synthetic Vision Systems Displays*, Proceedings of the AIAA Modeling and Simulation Technologies Conference, 2004