DOT/FAA/TC-23/66

Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport
New Jersey 08405

Operational Usability Assessment of the New Color Standard for Primary Terminal Air Traffic Control Displays

November 2023

Final report

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This report is available at the Federal Aviation Administration William J. Hughes Technical Center's Full-Text Technical Reports page: actlibrary.tc.faa.gov in Adobe Acrobat portable document format (PDF).

Technical Report Documentation Page

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

1. Report No. DOT/FAA/TC-23/66	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Operational Usability Assessment of the New Color Standard for Primary Terminal Air Traffic Control Displays		5. Report Date November 2023
		6. Performing Organization Code
7. Author(s) Todd R. Truitt, Ph.D., FAA and Brion Woroch, Ph.D., Diakon Solutions		8. Performing Organization Report No.
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address Federal Aviation Administration, Human Factors Division 800 Independence Avenue, S.W. Washington, DC 20591		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Objective: We emulated the new color standard, FAA HF-STD-010A, Baseline Requirements for Color Use in Air Traffic Control Displays, on a Terminal Radar Approach Control (TRACON) Primary Control Monitor (PCM). We then sought to identify human factors issues that may affect the broad implementation of the new color standard in existing and future air traffic control displays and identify areas where further research is required. **Method**: We conducted an operational usability assessment of the new color standard using a high-fidelity simulation platform, including dynamic air traffic scenarios. Six Air Traffic Control Specialists (ATCS) with normal color vision participated in three separate groups of two. The ATCS managed simulated traffic at two adjacent TRACON sectors using both the legacy and the new color standard. **Results**: We assessed airspace activity, communications, and subjective workload ratings, and found no significant differences between conditions or trials. The participants adjusted display brightness settings about twice as often with the new color standard but made these adjustments primarily for testing purposes rather than to improve legibility of a particular display element. Overall, the participants rated the legibility of objects on the PCM as moderate to high for both color standards. Participants reported that the new color standard provided consistent color-coding and significantly increased the legibility of target symbols, weather information, and lists. The participants also noted differences in their ability to reduce the brightness of some foreground elements, and the new colors of Air Traffic Proximity Alert cones and weather information. **Conclusion**: The objective data and ATCS participant responses did not identify any significant detrimental effects of the new color standard for ATCS with normal color vision. Rather, the participants preferred the new color standard in some instances. Restricting display element brightness and maintaining a minimum luminance contrast ratio are new concepts and practices for ATCS and may prevent them from using some established display strategies. Applying the new color standard fully to existing systems involves much more than simply replacing old color values with new ones and will require additional software development and testing on operational systems.

17. Key Words Color Use Standard User Interface Design Air Traffic Control Displays TRACON	18. Distribution Statement This document is available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia 22161. This document is also available from the Federal Aviation Administration William J. Hughes Technical Center at actlibrary.tc.faa.gov.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 72	22. Price

Contents

Figures

Tables

Acronyms

Acronym	Definition
ACE-IDS	ASOS Controller Equipment-Information Display System
ANOVA	Analysis of Variance
AOB	At or Below
ASOS	Automated Surface Observing System
ATC	Air Traffic Control
ATCS	Air Traffic Control Specialist
ATOP	Advanced Technologies and Oceanic Procedures
ATPA	Air Traffic Proximity Alert
CAMI	Civil Aerospace Medical Institute
CCR	Buchannan Field
CPC	Certified Professional Controller
CR1	Control Room 1
CVD	Color Vision Deficient
DCB	Display Control Bar
DESIREE	Distributed Environment for Simulation, Evaluation, and Experimentation
DRAT	Data Reduction and Analysis Tool
ER1	Experiment Room 1
ERAM	En Route Automation Modernization
FAA	Federal Aviation Administration
FPL	Full Performance Level
FL	Flight Level
HITL	Human-in-the-Loop
HWD	Hayward Executive Airport
IFR	Instrument Flight Rules
IRB	Institutional Review Board
MSL	Mean Sea Level
NAS	National Airspace System
NCT	Northern California TRACON
OAK	Oakland International Airport
PCM	Primary Control Monitor
PEQ	Post Experiment Questionnaire

PI	Primary Investigator
PMO	Program Management Office
PSQ	Post Scenario Questionnaire
RDHFL	Research, Development, and Human Factors Laboratory
SD	Standard Deviation
SECAP	Simulation Event Correlation and Event Processing
SFO	San Francisco International Airport
SJC	Norman Y. Mineta San Jose International Airport
SME	Subject Matter Expert
SQL	San Carlos Airport
STARS	Standard Terminal Automation Replacement System
TGF	Target Generation Facility
TRACON	Terminal Radar Approach Control
VFR	Visual Flight Rules
WAK	Workload Assessment Keypad

Executive summary

The Federal Aviation Administration (FAA) has published several human factors and design standards. An increased need for design standards in the development of air traffic control (ATC) displays arose as the prevalence of various information systems and displays increased in ATC facilities. Not only did the FAA need to know what information to display to Air Traffic Control Specialists (ATCS), but also how to display it effectively.

The FAA established a standard for color displays in 2007, FAA HF-STD-002, "Baseline requirements for color use in air traffic control displays" (Federal Aviation Administration, 2007) which established requirements for the legacy color standard that is used in existing ATC systems. Since that time, the FAA has conducted additional research on the use of color in ATC displays. With the increasing deployment of color displays in ATC facilities, the FAA recognized a need to develop color vision screening tests to ensure that ATCS could identify critical display information even though they may have had some form of color vision deficiency (Chidester, et al., 2011). Researchers revised the color vision screening tests over time to account for more complex, color-filled displays (Chidester, et al., 2013) that the FAA deployed like Ocean 21, Advanced Technology and Ocean Procedures (ATOP), and En Route Automation Modernization (ERAM).

The FAA continued to refine standards for the design of color displays and developed a standard color palette for use on ATC displays that would accommodate ATCS with color vision deficiencies (Gildea, Milburn, & Post, 2018). The standard color palette served as the foundation for the new FAA color standard, FAA HF-STD-010A, (Federal Aviation Administration, 2020) which supersedes the legacy standard, FAA HF-STD-002 (Federal Aviation Administration, 2007). In addition to defining a new color palette, the new color standard also requires maintaining minimum foreground-to-background luminance contrast ratios.

Researchers have tested the new color standard in static laboratory conditions (Gildea, Willems, Benincasa, Jack, & Post, 2020). However, ATC displays present complex and dynamic information that the FAA must assess in a realistic environment. ATC display information including images, icons, and alphanumeric characters are dynamic, move across the display, and sometimes overlap. To be effective, the color palette standard must ensure that information is presented in a manner and with color that is recognizable, discriminable, and legible for color-normal and color-vision deficient (CVD) ATCS.

In the current study, we implemented the new color standard, FAA HF-STD-010A, Baseline Requirements for Color Use in Air Traffic Control Displays (Federal Aviation Administration, 2020), on a Terminal Radar Approach Control (TRACON) Primary Control Monitor (PCM). We then compared the new color standard to the legacy color standard in use today to identify human factors issues that may affect the broad implementation of the new color standard in existing and future air traffic control displays, and identify areas where further research is required. Because the color palette was developed for ATCS with color vision deficiencies, we wanted to determine if the new color standard may affect ATCS with normal color vision. Furthermore, previous studies of the new color standard only used static displays or replays of air traffic events. This report documents an operational usability assessment of the new color standard using a high-fidelity simulation platform, including dynamic air traffic scenarios. Six ATCS with normal color vision participated in the usability assessment where they managed simulated traffic at two adjacent TRACON sectors using both the legacy and the new color standard.

We found no statistically significant differences between conditions or trials for airspace activity, communications, or ratings of subjective workload. The ATCS participants reported that the legibility of objects on the PCM was moderate to high for both color standards. Participants also said that the new color standard provided consistent color-coding and increased the legibility of target symbols, weather information, and lists. The participants noted differences in colors of the Air Traffic Proximity Alert cones, weather information, and their ability to reduce the brightness of some foreground elements.

The ATCS participants preferred the new color standard in some instances, and we did not identify any significant detrimental effects of the new color standard for ATCS with normal color vision. Fully implementing the new color standard will require additional software development and testing on operational systems. Furthermore, fully implementing the new color standard may prevent some ATCS from using established display strategies due to the restrictions imposed on display element brightness and the luminance contrast ratio.

The results of this operational usability assessment are based on a relatively small sample size that only provided statistical power to detect the largest of effects. There were also practical and technical limitations of the study that prevented us from testing every possible feature. However, based on our observations, objective data analyses, and feedback from the ATCS participants, we did not detect any significant issues with the new color standard that would prevent its implementation. Based on our findings, we recommend that the FAA provide ATCS with adequate training on the operation and use of the new color standard so they are able to effectively use and adapt display strategies. The FAA should also provide ATCS with adequate time to use in the new color standard in a simulation or training setting so they can develop and save new display preference sets for operational use. We also recommend further research on the design of weather information to determine the best method of weather presentation using the new weather colors. Finally, the FAA must consider the practical implications and challenges of implementing the new color standard on existing systems and software.

1 Introduction

This document presents the background, methodology, data analysis, results, and recommendations for an operational usability study of the new Federal Aviation Administration (FAA) color standard(Federal Aviation Administration, 2020). We designed the study to assess the new color standard as applied to the primary terminal Air Traffic Control (ATC) situation display, or primary control monitor (PCM). The FAA NextGen Human Factors Division (ANG-C1) funded this study in support of the FAA Air Traffic Organization Program Management Office – Program Control and Integration Team (AJM-131). Section 5 of this report provides a complete list of acknowledgements.

1.1 Background

In 2007, the FAA published standards document FAA HF-STD-002, "Baseline requirements for color use in air traffic control displays" (Federal Aviation Administration, 2007) which established requirements for the legacy color standard. Since that time, the FAA has implemented the color standard on new ATC displays and conducted various lines of research on the use of color in ATC displays. As the use of color on ATC displays became more prevalent, the FAA developed color vision screening tests to ensure that Air Traffic Control Specialist (ATCS) candidates could identify critical display information even though they may have had deficiency color vision in some form (Chidester, et al., 2011). Researchers continued to revise the color vision screening tests as more complex, color-filled displays such as those provided by Ocean 21, Advanced Technology and Ocean Procedures (ATOP), and En Route Automation Modernization (ERAM) platforms began to enter air traffic facilities (Chidester, et al., 2013).

The prevalence of color displays and the increased use of color for coding information on ATC displays led the FAA to develop a standard color palette (Gildea, Milburn, & Post, Development of a standard palette for color coding ATC displays (DOT/FAA/AM-18/22), 2018). The FAA Civil Aero Medical Institute (CAMI) developed the FAA color palette for color coding information on ATC displays in conjunction with Wright State University under a cooperative agreement. The new standard palette served as the foundation for the new FAA color standard, FAA HF-STD-010A, (Federal Aviation Administration, 2020) which supersedes the legacy standard, FAA HF-STD-002 (Federal Aviation Administration, 2007). In addition to defining a new color palette, the new color standard also requires maintaining minimum foreground-to-background luminance contrast ratios.

Researchers and user teams have vetted the new color palette in laboratory conditions (Gildea, Willems, Benincasa, Jack, & Post, 2020), but fielded ATC displays present images that are more complex and dynamic than could be achieved in previous studies. ATC display images, icons, and alphanumeric characters are dynamic and sometimes overlap. There are generally other untested, out-of-palette colors used for noncritical information that may cause confusion in discriminating and identifying information. To be effective, the color palette standard must ensure that colors are recognizable, discriminable, and legible for color-normal and color-vision deficient (CVD) ATCS. We encourage the Program Management Organization (PMO) and other FAA offices developing ATC operational systems to include the new color standard as a requirement for future ATC display design and evaluation.

1.2 Objective

Our objective was to identify any human factors considerations that would affect the broad implementation of the new color standard, "Baseline Requirements for Color Use in Air Traffic Control Displays," (Federal Aviation Administration, 2020) in existing and future ATC automation platforms. We conducted an operational usability assessment of the new color standard using a high-fidelity simulation platform, including dynamic air traffic scenarios. To support the usability assessment, we implemented the new color standard on the Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE) platform at the William J. Hughes Technical Center. The results of the current study provide a preliminary usability assessment of the new color standard, identify areas of concern, and suggest areas for further research.

1.3 Scope

The scope of the study was limited to the primary terminal ATC situation display, or PCM. The scope of the study was also limited to the implementation of the color standard on the PCM hardware (Section 2.2.1, Hardware) and emulation software (Section 2.2.2, Software) as described in this technical report. There may be issues associated with implementing the color standard on operational systems that we were not able to consider in this study.

1.4 Outcomes and benefits

Our long-term research goal is to provide empirical data that will inform the effective design of color ATC displays that accommodate both color-normal and CVD controllers. This report documents the results of an operational usability assessment study that compared the current color palette to a new color palette standard. The study identifies issues with the new color standard as applied to critical and non-critical information elements on the ATC display, proposes remedies as necessary, and informs future research. The results of the study will aid sponsors, stakeholders, and human factors practitioners in the development of standard ATC display designs. The human-in-the-loop (HITL) simulation platform will also serve as a foundation for the development and testing of alternative and future display designs.

2 Method

2.1 Participants

Six terminal ATCS participated in the study. The ATCS participants had experience at a Level 10 or above Terminal Radar Approach Control (TRACON) facility. All participants were FAA employees rated as Full Performance Level (FPL) Certified Professional Controllers (CPC). The participants had worked as an ATCS for an average of 20.75 years (*SD* = 4.39 years) and had controlled air traffic using the Standard Terminal Automation Replacement System (STARS) for an average of 9.08 years (*SD* = 4.94 years). Half of the participants had actively controlled traffic in a TRACON facility within the past year, while the other half had not. All participants had normal, or corrected to normal vison, and normal color vision.

The participants were subjected to minimal risk. Per the definition of minimal risk, the probabilities of harm or discomfort anticipated in this study were not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests. Throughout the study we maintained strict adherence to all federal, union, and ethical guidelines. All research personnel and participants were compliant with agency requirements for access to the research site and the safety and health procedures in effect at the time of data collection.

Participation in this study was voluntary and participants could withdraw at any time without penalty. The participants signed an informed consent statement prior to participation (see Appendix A). The principal investigator or another member of the research team was available to answer any participant questions regarding the study and any of the procedures involved.

2.2 Equipment

The study took place at the Research, Development, and Human Factors Laboratory (RDHFL) at the FAA William J. Hughes Technical Center. Researchers and participants primarily occupied Experiment Room 1 and Control Room 1 (ER1/CR1). Simulation pilots and their workstations were located in a separate room near ER1/CR1.

2.2.1 Hardware

We constructed two Terminal workstations (Figure 1) in ER1. Each workstation included a Barco 2K PCM. The PCM, designed specifically for ATC, provided a resolution of 2048 x 2048 pixels and viewable display of 19.83” x 19.83” (503.8 mm x 503.8 mm) and 28.05” (712.4 mm) diagonal. Each workstation also included an emulation of the Automated Surface Observing System (ASOS) Controller Equipment-Information Display System (ACE-IDS), a Standard Terminal Automation Replacement System (STARS) keyboard and trackball, and an emulated terminal Voice Switching and Communication System (VSCS). The ACE-IDS used a 24” touchscreen with an active display area of 12” (304.8 mm) x 21” (533.4 mm) with a 1900 x 1080-pixel resolution. Simulation pilots and SMEs used workstations to affect simulated aircraft movements and communications. Each simulation pilot workstation included a computer, keyboard, mouse, display of aircraft information, and communications system. A Workload Assessment Keypad (WAK; Stein, 1985) and Easy Button were located at each workstation. Ceiling-mounted color video cameras were located above and behind each participant workstation. The room was dimly light with overhead fluorescent lights turned off and dim backlights illuminated to prevent glare on the PCMs. We used a Photo Research PR-788 spectroradiometer to measure the PCMs and generate display characteristic files. We also used a Photo Research Inc. PR-524 LiteMate photometer to measure ambient lighting levels. The PR-788 and PR-524 devices were both in calibration at the time of our measurements.

Figure 1. Terminal ATCS workstations in the RDHFL

2.2.2 Software

The simulation used the DESIREE along with the Target Generation Facility (TGF). DESIREE emulated the current baseline STARS interface functionality. The TGF provided aircraft performance models, generated aircraft tracks based on pre-defined flight plans, and enabled the simulation pilot workstations. Both DESIREE and TGF provided data collection capabilities (e.g., ATCS display settings; number and type of controller keyboard entries; number of aircraft maneuvers; time and distance flown within each sector).We used the TGF Data Reduction and Analysis Tool (DRAT) along with the Simulation Event Correlation and Event Processing (SECAP) tool to calculate objective measures. We used FFmpeg to record audio and video from the PCM.

2.2.3 Implementing the FAA color standard

The DESIREE emulated the STARS baseline functionality, user interface, and implemented the appropriate color standard. We implemented the new FAA color standard (Federal Aviation Administration, 2020) on DESIREE by replacing the legacy STARS colors in DESIREE with corresponding color values from the new color standard. We defined the weather as a background element and all other elements were foreground elements, with the exception of the weather stipples (levels of weather) and history trails. Once we defined the foreground and background elements, we created an algorithm that enforced the 3:1 foreground-to-background luminance contrast ratio by restricting the possible brightness ranges for each adjustable display element. For example, the brightness ranges for a given foreground display element (e.g., position symbol) varied based on the brightness of the background display elements (i.e., weather) and vice versa. Therefore, the brightness ranges of foreground and background elements were dependent upon one another and changed based on personal preferences. ATCS participants could adjust various individual display settings either via the software display control bar (DCB) or physical knobs on the display control panels. As shown in Figure 1, the DCB is located at the top of the PCM and the display control panels are on either side of the PCM.

We designed the details of the weather display for the new color standard based on feedback from ATC SMEs and rapid prototyping activities. A new design for the weather presentation was required because there are a different number of background weather colors in the legacy and new color standards. The legacy color standard presents six levels of weather by combining two background colors with three stipple patterns (none, sparse, dense). However, the new color standard provides three background weather colors and no defined stipple pattern. We presented the six levels of weather using the new color standard by combining the three background weather colors with two stipple patterns (none, sparse). We kept the color of the weather stipples consistent with the legacy colors and replaced it with a corresponding color (gray) from the new color standard. Figure 2 and Figure 3, respectively, show the legacy weather presentation and the new weather presentation based on the new color standard that we implemented.

The weather stipples and history trails were not subject to the same brightness restrictions as other elements because their assigned colors would easily violate the luminance contrast enforcement algorithm. A light-colored (grey) background element or dark-colored foreground element would have violated the enforcement algorithm and prevented the display of information. We also treated the brightness settings for data blocks as a special case because of the dwell feature. When the user enables the dwell feature and places the trackball pointer over a data block, the data block brightness increases by 20%. We restricted the user-adjusted brightness of the data blocks to 80% maximum so that with dwell enabled, the data block brightness increased to the maximum restricted brightness.

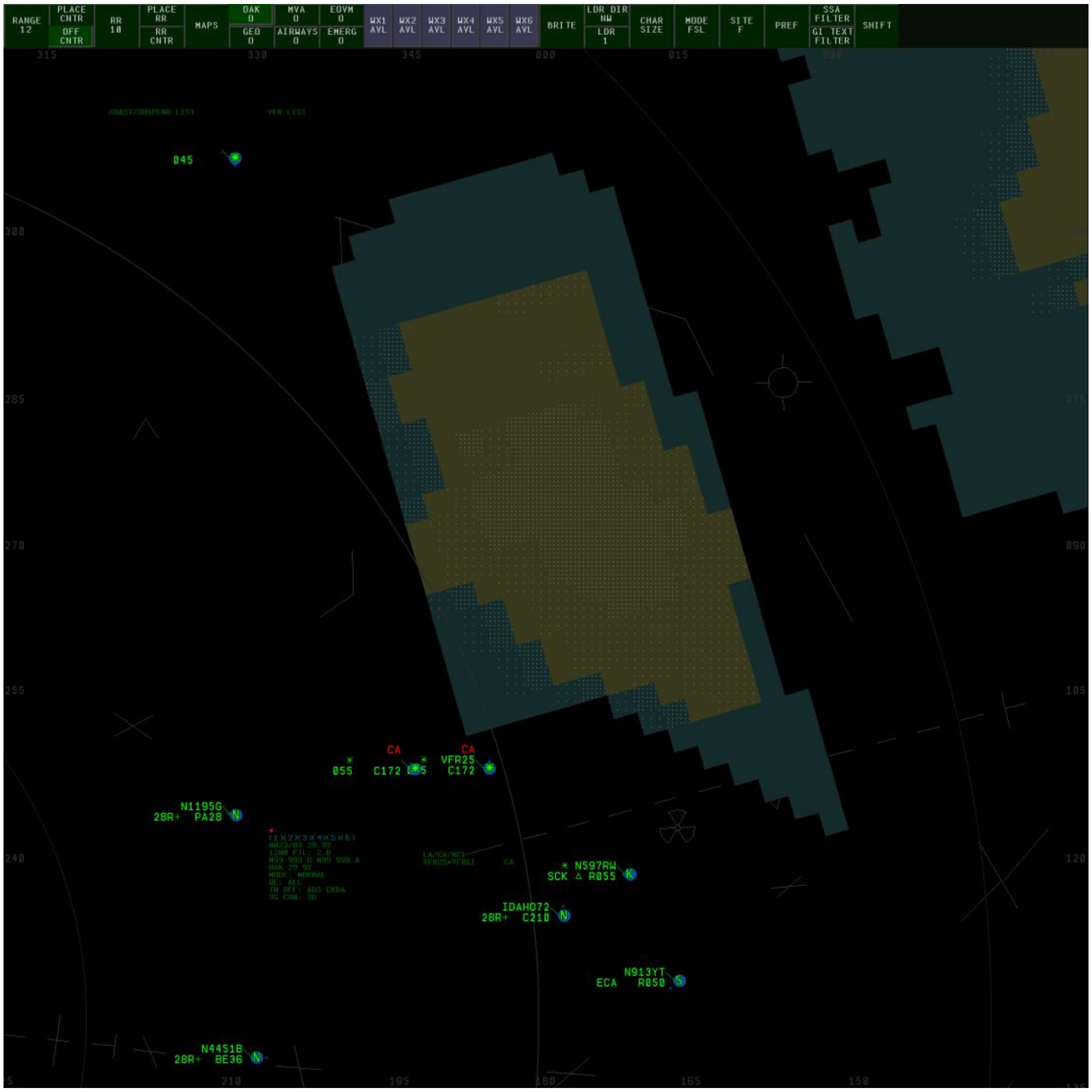

Figure 2. Legacy weather presentation on a PCM

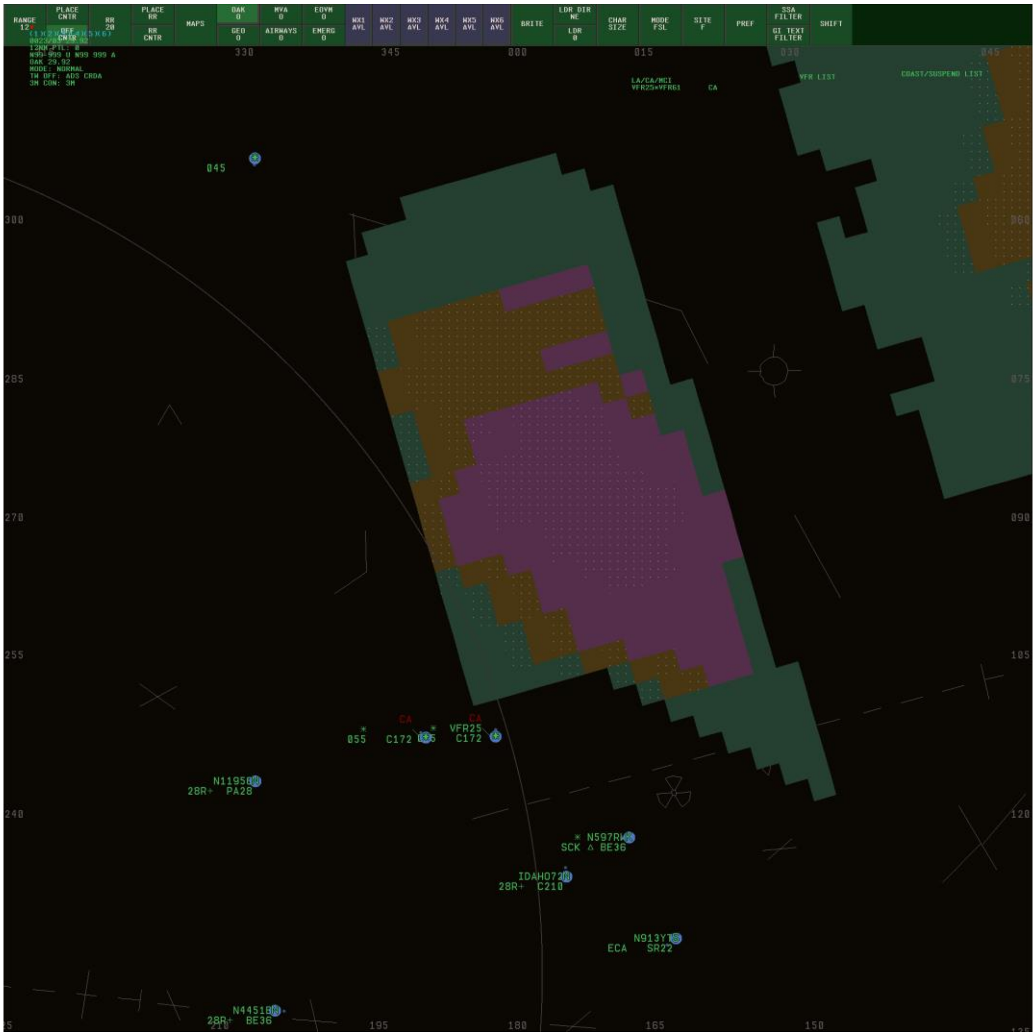

Figure 3. New weather presentation on a PCM

Prior to the study, we used a calibrated Photo Research PR-788 spectroradiometer to measure each display and generate display characteristic files. The DESIREE software then adjusted the output RGB values for each display to ensure that the displays presented the exact same RGB values to the user for a known brightness configuration.

2.3 Materials

2.3.1 Informed consent

Each participant read and signed an informed consent statement before the experiment (see Appendix A). The informed consent statement described the study and the rights and responsibilities of the participants, including that their participation was voluntary, all information they provide is anonymous and confidential, and indicated any foreseeable risks to which they may be subjecting themselves. Signing the form indicated their voluntary consent to participate in the study.

All information that the participants provided is anonymous. Researchers attached a participant code to all data for research purposes. We will not release the participants' names or identities in any reports. All data collected in the study was for scientific purposes only and researchers must keep the data confidential by law. Laboratory personnel will not disclose or release any Personally Identifiable Information to any FAA personnel or elsewhere, or publish it in any report, except as may be required by statute.

2.3.2 Questionnaires

We used a Background Questionnaire (see Appendix B) to collect general background information about each participant and to assess their prior level of experience. The participants completed a Post Scenario Questionnaire (PSQ; see Appendix C) to provide feedback after each scenario. At the end of the study, the participants completed the Post Experiment Questionnaire (PEQ; see Appendix D) to provide feedback on the overall study.

2.3.3 Airspace

We used airspace comprising modified Mulford and Grove sectors of Northern California TRACON (NCT). We removed the complex altitude shelf-structure of the Mulford and Grove sectors to simplify operations within the sectors. We also simplified the surrounding airspace so we could recruit participants from TRACON facilities across the NAS and so participants could learn the airspace in a relatively short period of time. We consolidated the multiple sectors that surround Mulford and Grove into North and South sectors to reduce the number of sector handoff symbols and radio frequencies that participants had to memorize. Figure 4 depicts the simplified airspace.

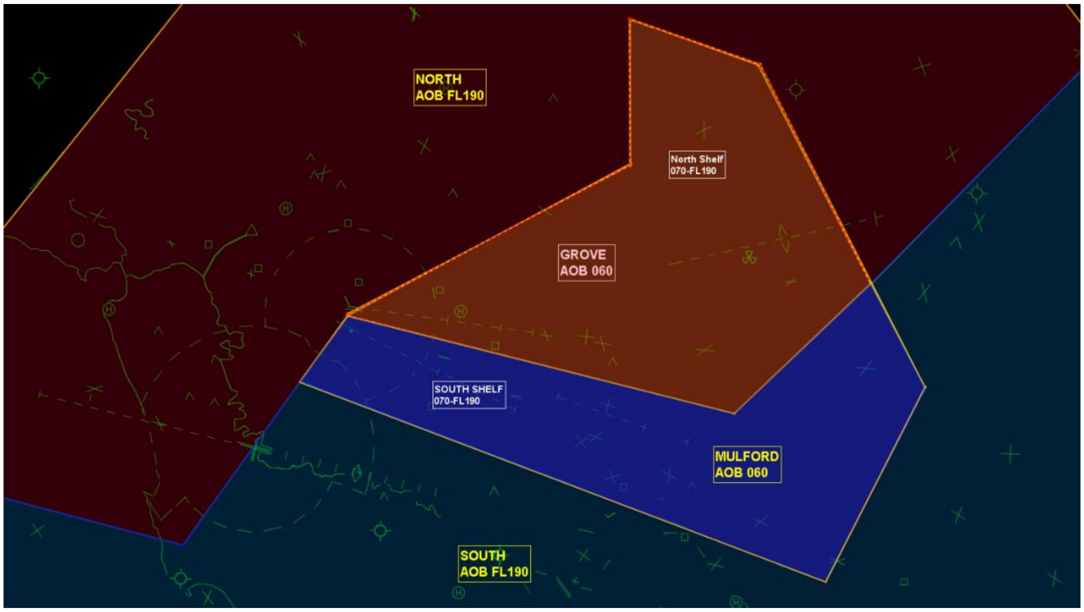

Figure 4. Simplified Mulford, Grove, North, and South sectors with altitude shelves

The participants controlled traffic in the Grove and Mulford sectors and managed arrivals into Oakland International Airport (OAK). As Figure 5 shows, OAK comprises four runways: 30/12, 28L/10R, 28R/10L, and 33/15. We simulated air traffic for a “West” configuration that required arrivals to use runways 30, 28L, and 28R. We did not use runway 33/15. The Grove sector included airspace at or below (AOB) 6,000 ft mean sea level (MSL). The Grove sector was located above the final approach course to OAK runways 28L and 28R and was responsible for directing traffic to these runways. The Mulford sector included airspace AOB 6,000 ft. The Mulford sector was located above the final approach course to OAK runway 30 and the final approach course to Hayward Executive Airport (HWD) runway 28L and was responsible for directing traffic to those runways.

The North and South sectors were ghost sectors. An ATC SME or simulation pilot staffed each ghost sector. The North sector managed traffic AOB 19,000 ft MSL (FL190) and from between 7,000 ft MSL to FL190 over the Grove sector. The South sector managed traffic AOB FL190 and from between 7,000 ft MSL to FL190 over the Mulford sector.

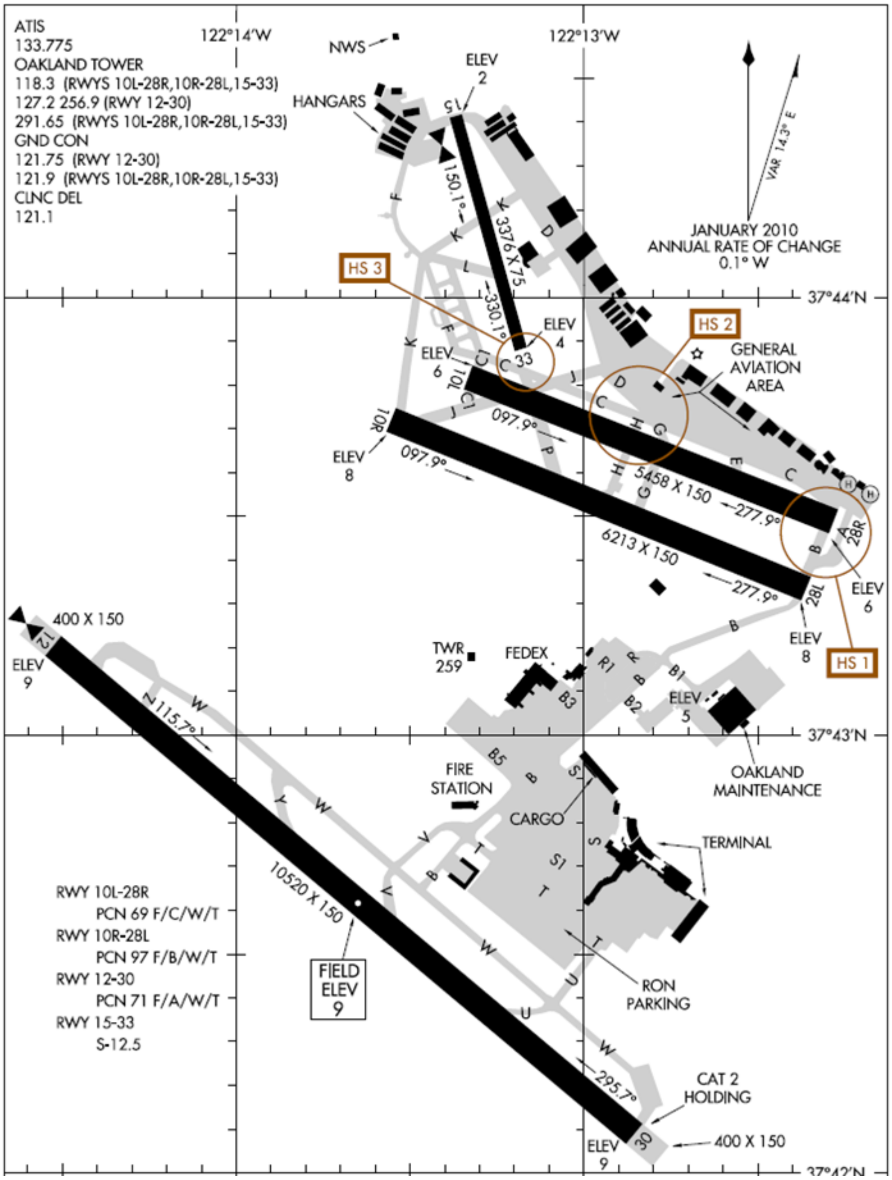

Figure 5. Airport diagram of Oakland International Airport (OAK)

2.3.4 Air traffic scenarios

We repurposed air traffic scenarios from a previous study (Truitt, Zingale, & Konkel, 2016) to reduce scenario development time and ensure realistic traffic patterns. There were 87 total aircraft in the base scenario (61 arrivals, 19 departures, and 7 overflights). Eighteen of the aircraft were visual flight rules (VFR) flights, and 69 of the aircraft were instrument flight rules (IFR) flights. The scenario had a total of 17 arrivals at OAK 30, 12 arrivals and 2 departures at OAK 28R, 6 arrivals and 4 departures at HWD 28L. The scenario also contained other traffic that impacted the participants' airspace. There were 13 arrivals at San Francisco International

Airport (SFO) 28R, 13 arrivals at SFO 28L, and 2 departures at SFO 28R. Of the 19 departures, there were 6 from Norman Y. Mineta San Jose International Airport (SJC), 4 from Buchanan Field (CCR), and 1 from San Carlos Airport (SQL). We created 12 versions of the base scenario by changing only the call signs of the aircraft. We used four versions of the scenario for the training runs and the remaining eight versions for the test runs. The small number of participants and test runs in this usability assessment precluded a fully counterbalanced experimental design. Therefore, we decided to use the same traffic patterns and levels in all test runs to eliminate a scenario confound while still controlling for order effects of condition (legacy color standard vs. new color standard) to the extent possible.

We also added a dynamic weather pattern to the base scenario that replicated in every scenario version. The weather pattern was located near primary air traffic routes to create some overlap between the weather pattern and aircraft information (e.g., position symbols, data blocks) without interfering in the overall scenario. Figure 1 shows the location of weather, which slowly changed and moved South during the scenario. We instructed the participants that although weather was present, VFR conditions prevailed, and the weather did not cause participants to hold aircraft outside of the sector or implement rerouting.

2.4 Pre-testing

We conducted a shakedown of the study prior to formal data collection to test all simulation and data collection capabilities and to test and refine the experiment procedure. The shakedown also served as an opportunity to familiarize simulation pilots with the airspace and procedures. The shakedown relied on the research team, including experimenters, software and hardware engineers, and SMEs. During the shakedown, we executed the experiment procedure as planned while monitoring all data collection and storage. The research team documented, corrected, and re-tested any substantial issues prior to conducting the study.

2.5 Experimental design

We designed the current study to compare the legacy color standard to the new color standard using a 2 (Color Standard – New v. Legacy) X 2 (Trial – First v. Second) within-subjects design. Three groups of two ATCS participated (N =6). We were unable to counterbalance conditions because of the small sample size and number of trials and note the low statistical power to detect any significant differences.

2.6 Procedure

2.6.1 Schedule of events

Each group of participants travelled to the RDHFL on a Monday, participated in the study Tuesday-Thursday, and then returned to their facilities on Friday. Table 1 contains the notional daily schedule for each group of participants. We conducted an in briefing and training on Day 1 (Tuesday) and then conducted data collection on Days 2 and 3 (Wednesday and Thursday).

Table 1. Notional daily schedule

Time	**Day 1 - Tuesday**	**Time**	**Day 2 - Wednesday**	**Time**	**Day 3 - Thursday**
0830	Informed Consent & In Brief	0830	Training as Needed	0830	Prep as Needed
0900	Airspace & DESIREE Training	0900	Test Scenario 1	0900	Test Scenario 7
1000	Break	0945	Break	0945	Break
1015	Airspace & DESIREE Training	1000	Test Scenario 2	1000	Test Scenario 8
1145	Break	1045	Break	1045	Break
1245	Practice Scenario 1	1100	Test Scenario 3	1100	Test Scenario Make Up
1330	Break	1145	Break	1145	Post Experiment Survey
1345	Practice Scenario 2	1245	Test Scenario 4	1200	Break
1430	Break	1330	Break	1330	Caucus
1445	Practice Scenario 3	1345	Test Scenario 5	1345	Survey Review
1530	Break	1430	Break	1430	Break
1545	Practice Scenario 4	1445	Test Scenario 6	1445	Summary & Discussion
1630	End of Day	1530	Break	1530	Break
		1545	Discussion as needed	1545	Out Brief
		1630	End of Day	1630	End of Day

2.6.2 Data collection

Once the participants arrived at the RDHFL, a Principal Investigator (PI) briefed them on the background and objectives of the study. The PI also provided details about the daily schedule (see Table 1) and general laboratory procedures. After the in briefing, each participant and a PI signed an Informed Consent Statement (see Appendix A) and the participants completed the Biographical Questionnaire (see Appendix B). Next, an ATC SME presented an overview of the airspace, procedures, and laboratory environment including the functionality of the TRACON emulation, DESIREE.

After the classroom training, the participants began hands-on training. The participants controlled air traffic in four practice scenarios (two scenarios at each sector, Mulford and Grove). Each practice scenario was 30 minutes in duration and used the legacy color standard. Before each practice run, an experimenter assigned each participant to either the Mulford or Grove sector and adjusted the room lighting. The participants prepared their workstation and saved PCM brightness and adjustment preference settings for later use. An experimenter then read the WAK instructions out loud to the participants. We instructed participants to respond to the WAK during practice scenarios to familiarize them with the WAK device and procedure. Once the participants were ready, we instructed them to control traffic normally as they would in the field and began the scenario. The participants controlled air traffic at their assigned sector for one scenario and then switched sectors. This process repeated until each participant completed two practice scenarios at each sector. The ATC SME was available throughout the training scenarios to answers any questions from the participants.

After completing the practice scenarios, the participants controlled air traffic in a total of eight data collection scenarios. The data collection scenarios were 30 minutes in duration. The researchers assigned each participant to one of the two sectors, Mulford or Grove, and adjusted the room lighting. Sometimes the participants did not switch sectors after a scenario, depending on the counterbalancing scheme. We partially counterbalanced the order of conditions across participant groups as shown in Table 2. Before each scenario, the participants had the opportunity to prepare their workstation and adjust the ambient task lighting, keyboard brightness, and all PCM settings including brightness. Once the participants were comfortable, we used a Photo Research Inc. PR-524 LiteMate photometer to measure ambient lighting levels at each participant’s workstation. We measured illumination at the location between the participant’s keyboard and trackball which were directly in front of the PCM (see Figure 1). After measuring work surface illumination, a PI instructed the participants to press the Easy Button at any time during a scenario to indicate a point for further review and then read the

WAK instructions (see Appendix E) to the participants out loud. We read the WAK instructions to the participants before each run to remind them of the importance of the measure and to reinforce the rating scale. Once the participants were ready, we instructed them to control traffic normally as they would in the field, and began the scenario. During each scenario, the participants were responsible for controlling air traffic, communicating, coordinating, and maintaining flight data. Every 2 min, the WAK prompted the participants for a subjective workload rating by emitting an alerting sound (a high-pitched chirp) and illuminating the WAK buttons. Each participant had 30 s to respond to the prompt by selecting one of the ten numbered WAK buttons to indicate their current level of workload.

DESIREE recorded each participant's PCM settings at the beginning of each scenario. DESIREE also recorded any changes to a participant's PCM settings during a scenario. When the participants controlled traffic in the new color standard condition, we required them to use PCM brightness and display settings that conformed to the standard's requirements. During runs with the new color standard, DESIREE prevented the participants from adjusting the PCM settings outside the specifications as described in FAA HF-STD-010A (Federal Aviation Administration, 2020). At the completion of each scenario, a researcher turned on the overhead lights in ER1 and the participants completed the PSQ (see Appendix C).

During each scenario, we recorded digital audio and video data from each TRACON workstation VSCS and PCM using FFmpg. We recorded the controller and pilot transmissions via the VSCS including controller conversations via the headset microphones. Cameras mounted on the ceiling recorded an over-the-shoulder view of each control position.

Table 2. Counterbalancing order by group, sector position, and condition

Group	**Run**	**Mulford**	**Grove**	**Condition**
1	1	P1	P2	Legacy
	2	P1	P2	New
	3	P2	P1	Legacy
	4	P2	P1	New
	5	P1	P2	New
	6	P2	P1	Legacy
	7	P2	P1	New
	8	P1	P2	Legacy
2	1	P4	P3	Legacy
	2	P4	P3	New
	3	P3	P4	Legacy
	4	P3	P4	New
	5	P4	P3	New
	6	P3	P4	Legacy
	7	P3	P4	New
	8	P4	P3	Legacy
3	1	P6	P5	New
	2	P6	P5	Legacy
	3	P5	P6	New
	4	P5	P6	Legacy
	5	P6	P5	New
	6	P5	P6	Legacy
	7	P5	P6	New
	8	P6	P5	Legacy

3 Results & discussion

The purpose of the data analysis was to identify usability issues with the new color standard. We determined results of inferential statistical tests to be significant when *p* values were less than or equal to .05. We report inferential statistics and effect sizes (e.g., Cohen's *d* for *t* tests and partial eta-squared ($\eta_p{}^2$) for ANOVAs) for all statistically significant results[1]. To account for the circularity assumption that may be violated with small sample sizes (Hays, 1988) and (Kirk, 1982), we analyzed data that produced significant ANOVA results with the Geisser-Greenhouse (G-G) *F* test (Geisser & Greenhouse, 1958). A significant G-G *F* test indicated a highly significant result. A non-significant G-G *F* test indicated a violation of the circularity assumption and then we then applied the Box adjustment – Huynh-Feldt (H-F) *F* test or adjusted *F* test (Huynh & Feldt, 1970). The H-F *F* test was the final determinant of whether a significant effect was present or not. We present summary and descriptive statistics for the Biographical questionnaire, PSQ, PEQ and relevant dependent measures collected by DESIREE and TGF.

3.1 Airspace activity

We recorded measures associated with aircraft activity to assess the participants' task load and airspace efficiency in each condition. If the color standards affected airspace efficiency or the participants' ability to control air traffic in some way, it may be possible to detect that difference in airspace activity. We analyzed each of the airspace activity measures (number of aircraft handled, distance flown, and time in sector) for each sector (Mulford and Grove) to detect any differences between the legacy and new color standard. We used a 2 (Condition – Legacy Color Standard vs. New Color Standard) X 2 (Trial – First vs. Second) repeated measures Analysis of Variance (ANOVA) to analyze the data set while accounting for circularity assumptions as appropriate. We analyzed the airspace activity data for each sector separately because there were different traffic levels and patterns in each sector that could have affected these measures. We only analyzed and report data for aircraft that the participants controlled and traversed either the Mulford or Grove geographical sectors.

3.1.1 Number of aircraft handled

We recorded the total number of controlled IFR aircraft that the participants handled in each scenario to gauge the participants' task load and ensure that task load was the same in both conditions (see Table 3). On average, participants handled about 34 aircraft in the Mulford sector

---

[1] Cohen (1988, 1992) describes the use of Cohen's *d* and partial eta squared to evaluate effect size. For both measures, a value of 0.20 is considered a small effect, 0.50 is considered a medium effect, and 0.80 or higher is considered a large effect.

and about 39 aircraft in the Grove sector during each 30-minute scenario. The number of aircraft handled and the associated task load within each sector was comparable in both color standard conditions for Mulford (F (1, 5) = 0.64) and Grove (F (1, 5) = 0.14) sectors. Likewise, there was no difference between trials for the number of aircraft handled in either the Mulford (F (1, 5) = 0.01) or Grove (F (1, 5) = 1.37) sectors.

Table 3. Mean number of aircraft handled by sector, condition, and trial

	Mulford ***M (SD)***	**Grove** ***M (SD)***
Legacy Color Standard	33.4 (4.4)	39.7 (4.1)
New Color Standard	34.5 (2.7)	39.3 (2.2)
Trial 1	33.9 (3.2)	38.9 (3.6)
Trial 2	34.0 (2.1)	40.1 (2.4)

3.1.2 Distance flown in sector

We recorded the total distance flown by aircraft during each scenario to assess potential changes in airspace efficiency that may be due to the color standard condition (see Table 4). Aircraft flew the same distance in each sector regardless of condition or trial. We found that the new color standard did not have a measurable effect on total distance flown for either the Mulford (F (1, 5) = 0.15) or Grove (F (1, 5) = 0.13) sectors. Distance flown did not differ significantly between trials for either the Mulford (F (1, 5) = 0.72) or Grove (F (1, 5) = 0.09) sector.

Table 4. Mean total distance flown (nm) by sector, condition, and trial

	Mulford ***M (SD)***	**Grove** ***M (SD)***
Legacy Color Standard	360.9 (41.6)	339.9 (57.7)
New Color Standard	364.7 (35.8)	336.0 (51.9)
Trial 1	359.7 (42.3)	336.3 (58.1)
Trial 2	365.9 (29.0)	339.6 (58.1)

3.1.3 Time in sector

We recorded the time flown inside the Mulford and Grove geographical sectors during each scenario. There was a lot of variability in the distance flown by each aircraft due to the number of unique flight plans in the scenario, and so aircraft time in sector exhibited a large amount of variability. The new color standard did not significantly affect the average time that aircraft flew in the Mulford (F (1,5) = 0.86) or Grove (F (1,5) = 0.25) sectors. The average time that aircraft flew in a sector was similar across trials for both the Mulford (F (1, 5) = 0.28) and Grove (F (1, 5) = 0.36) sectors. Table 5 shows the mean time in sector (s) for Mulford and Grove.

Table 5. Mean time in sector (s) by sector, condition, and trial

	Mulford ***M (SD)***	**Grove** ***M (SD)***
Legacy Color Standard	6438.5 (698.2)	7685.6 (1156.8)
New Color Standard	6645.5 (646.9)	7555.8 (1146.8)
Trial 1	6500.6 (703.1)	7568.6 (1070.6)
Trial 2	6583.4 (452.3)	7672.8 (1182.6)

3.2 Radio communications

We recorded the number and duration of controller-to-pilot push-to-talk (PTT) radio transmissions to assess the communication task load of controllers and to see if there was any difference between the color standard conditions.

We analyzed the number and duration of PTT transmissions for each sector (Mulford and Grove) to detect any differences between the legacy and new color standard using a 2 (Condition – Legacy Color Standard vs. New Color Standard) X 2 (Trial – First vs. Second) repeated measures Analysis of Variance (ANOVA) and accounted for circularity assumptions as appropriate. We analyzed the PTT data for each sector separately because there were different traffic levels and patterns in each sector that could have affected PTT transmissions.

Neither the mean number nor duration of controller-to-pilot PTT transmissions differed significantly between condition or trial for the Mulford or Grove sector. For the Mulford sectors, PTT transmissions were similar in number (F (1, 5) = 0.001) and duration (F (1, 5) = 0.73) regardless of the color standard condition. Likewise, the number (F (1, 5) = 0.51) and duration (F (1, 5) = 1.25) of PTT transmissions in the Mulford sector did not differ across trials. Results were similar for the Grove sector where similar number (F (1, 5) = 0.17) and duration (F (1, 5) = 0.91) of PTT transmissions between conditions, and a similar number (F (1, 5) = 0.003) and duration (F (1, 5) = 1.92) of PTT transmissions between trials. Table 6 and Table 7, respectively, show the mean number and duration of PTT transmissions for each sector by Condition and Trial.

Table 6. Mean number of controller-to-pilot PTT transmissions by sector, condition, and trial

	Mulford ***M (SD)***	**Grove** ***M (SD)***
Legacy Color Standard	113.92 (16.79)	98.75 (26.63)
New Color Standard	114.08 (12.72)	100.25 (25.33)
Trial 1	112.67 (13.96)	99.58 (28.12)
Trial 2	115.33 (15.18)	99.42 (23.35)

Table 7. Mean duration (sec) of controller-to-pilot PTT transmissions by sector, condition, and trial

	Mulford ***M (SD)***	**Grove** ***M (SD)***
Legacy Color Standard	4.12 (0.50)	4.11 (0.50)
New Color Standard	4.06 (0.51)	3.98 (0.56)
Trial 1	4.14 (0.49)	3.98 (0.54)
Trial 2	4.04 (0.52)	4.10 (0.49)

3.3 Subjective workload

We analyzed the WAK ratings and response times for each sector (Mulford and Grove) to detect any differences in subjective workload between the legacy and new color standard using a 2 (Condition – Legacy Color Standard vs. New Color Standard) X 2 (Trial – First vs. Second) X 15 (Time Interval) repeated measures ANOVA. We analyzed the WAK data for each sector separately because there were different traffic levels and patterns in each sector that could have affected subjective ratings of workload. We replaced any missing data with the median rating from the respective trial.

Neither the participants' WAK ratings nor response times differed significantly between conditions or trials for the Mulford or Grove sector (see Table 8 and Table 9). For the Mulford sector, the participants' WAK ratings did not differ significantly between conditions (F (1, 5) = 2.24) or trials (F (1, 5) = 003). Likewise, the participants' WAK response time was similar between conditions (F (1, 5) = 1.06) and trials (F (1, 5) = 3.72). The participants WAK ratings at the Grove sector did not differ significantly between conditions (F (1, 5) = 0.02) or trials (F (1, 5) = 2.79), and their WAK response times were also similar between conditions (F (1, 5) = 0.16) and trials (F (1, 5) = 1.89). The participants' WAK ratings did change significantly over time in both the Mulford ($F(14, 70) = 10.65$, $p < .001$, $\eta p^2 = 0.68$) and Grove ($F(14, 70) = 7.86$, $p < .001$, $\eta p^2 = 0.61$). Figure 6 and Figure 7, respectively, show the mean WAK responses for each sector by condition and time interval.

Table 8. Mean WAK ratings by sector, condition, and trial

	Mulford ***M (SD)***	**Grove** ***M (SD)***
Legacy Color Standard	2.21 (1.06)	2.34 (1.18)
New Color Standard	2.33 (0.95)	2.34 (1.22)
Trial 1	2.27 (0.80)	2.21 (1.18)
Trial 2	2.26 (1.22)	2.47 (1.25)

Table 9. Mean WAK response time (s) by sector, condition, and trial

	Mulford ***M (SD)***	**Grove** ***M (SD)***
Legacy Color Standard	2.63 (0.37)	2.48 (0.43)
New Color Standard	2.44 (0.11)	2.43 (0.48)
Trial 1	2.66 (0.18)	2.34 (0.29)
Trial 2	2.41 (0.24)	2.56 (0.60)

The participants' WAK ratings and response times indicate that the new color standard did not significantly affect workload. Overall, the participants rated workload during the scenarios as low. Although the participants rated workload as changing over time, there was no evidence to suggest these changes in workload were due to the color standard that the participants used.

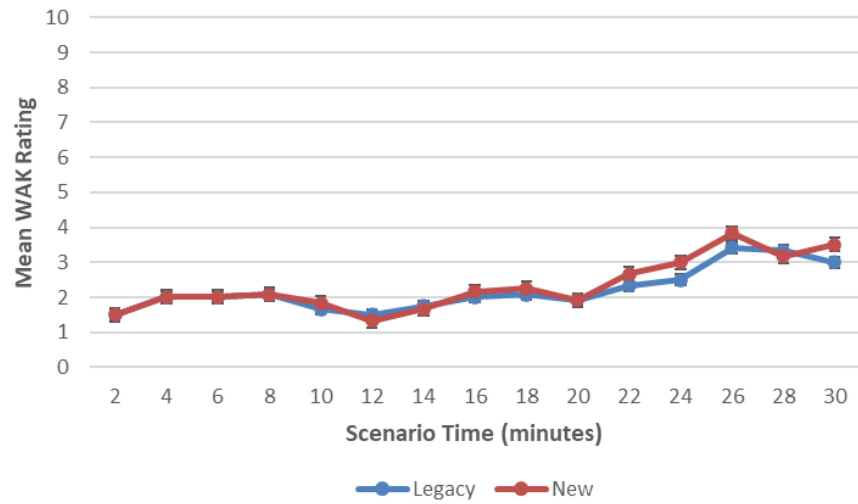

Figure 6. Mean WAK rating for Mulford sector by condition and interval

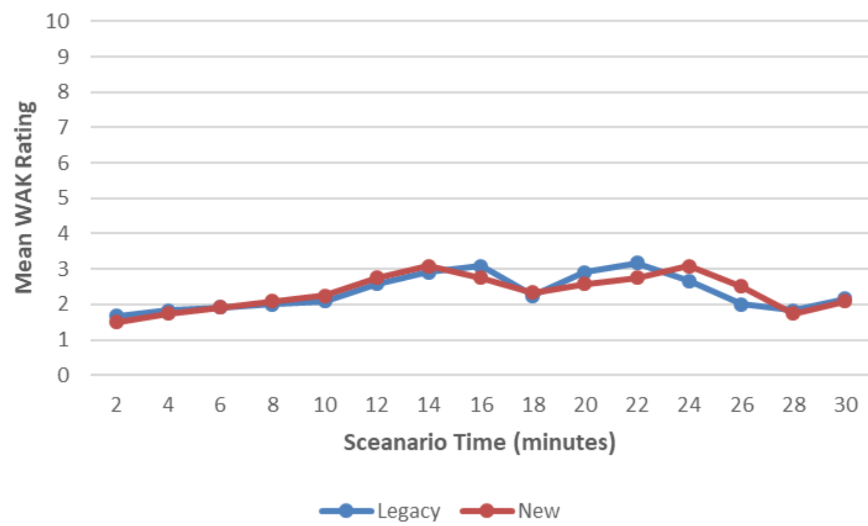

Figure 7. Mean WAK rating for Grove sector by condition and interval

3.4 Display brightness adjustments

We recorded the type, number, and duration of display brightness adjustments that the participants made during each scenario to understand how the participants changed their PCM while they controlled air traffic. We also recorded the method that participants used to make brightness adjustments - either via the software Display Control Bar (DCB) or physical knobs on the display control panel. We only counted brightness adjustments that persisted for 1 s or longer to reduce the chance of counting interim adjustment values. Recording display brightness adjustments on the PCM provided an objective measure of how the participants interacted with the display elements when they used each color standard and could help us identify usability issues.

The participants loaded saved preference sets and made most of the adjustments to their displays and workstations before each run. The participants made relatively few display adjustments while they controlled the simulated air traffic. Somewhat surprisingly, the participants only made display brightness adjustments in 17 of the 48 total runs between both sectors, Mulford and Grove (3 groups X 8 data collection runs X 2 sectors).

We recorded a total of 153 display adjustments made during the data collection scenarios, an overall average of 6.4 adjustments per 30-minute scenario. The participants made 53 adjustments when they used the legacy color standard, compared to 100 adjustments when they used the new color standard. Overall, the participants made nearly twice as many display brightness adjustments when they used the new color standard compared to the legacy color standard, regardless of sector. Table 10 shows the total number of display element adjustments by sector and condition.

We noted the number of adjustments for the range rings and tools settings that only occurred at the Mulford sector with the new color standard. Similarly, the participants only adjusted Map A and Other when they used the new color standard. The participants most often adjusted limited data blocks, weather, and range rings. Based on our observations, discussions with participants, and participant comments, the overall pattern of display setting adjustments suggests that the participants were testing the limits of the new color standard and identifying issues that they could later rate and comment on using the PSQ. As we note in the subsequent PSQ results, the participants also identified several artifacts that resulted from our implementation of the new color standard on DESIREE. The participants actively searched for the limits of usability and made display element adjustments during the scenarios to identify and evaluate those usability issues as well as other artifacts that were present.

3.5 Questionnaire ratings and responses

The participants responded to the PSQ (Appendix C) after each scenario and completed the PEQ (Appendix D) at the end of the study. Each questionnaire provided the participants an opportunity to respond to open-ended questions and provide subjective ratings and additional comments.

3.5.1 Post scenario questionnaire

The PSQ used a 7-point Likert scale. For PSQ items 1 through 5, a rating of “1” represented “Very Negatively” and a rating of “7” represented “Very Positively”. For PSQ items 6 through 31, a rating of “1” represented “Very Low” and a rating of “7 represented “Very High”. We analyzed the participants’ PSQ ratings using the non-parametric Wilcoxon Signed-Rank Test (Wilcoxon, 1945). Table 11 presents the median ratings and interquartile ranges for each PSQ item.

Table 10. Total number of display element adjustments by sector and condition

	Mulford		Grove		Total	
Display Element	**Legacy**	**New**	**Legacy**	**New**	**Legacy**	**New**
BCN – Beacon Code	0	0	0	3	0	3
CMP – Compass Rose	0	0	0	2	0	2
DCB – Display Control Bar	2	0	0	0	2	0
FDB – Full Data Block	3	4	2	2	5	6
HST – History Trails	0	0	0	3	0	3
LDB – Limited Data Block	13	5	4	4	17	9
LST - Lists	2	2	2	4	4	6
MPA – Map A	0	4	2	4	2	8
MPB – Map B	0	2	0	2	0	4
OTH - Other	0	5	0	3	0	8
POS – Position Symbol	3	0	0	2	3	2
PRI – Primary Target	1	0	0	2	1	2
RR – Range Rings	0	13	2	6	2	19
TLS - Tools	0	12	0	0	0	12
WX - Weather	8	6	7	3	15	9
WXC – Weather Contrast (Stipple)	0	4	2	3	2	7
Total	**32**	**57**	**21**	**43**	**53**	**100**

Overall, the participants indicated on the PSQ that neither the legacy nor new color standard impacted the safety or efficiency of operations, awareness of potential conflicts, ability to coordinate with other sectors, or workload. The participants rated the legibility of objects on the PCM as moderate to high for both color standards (median ratings ranged between 4 and 7). However, the participants thought that the new color standard provided greater legibility for target symbols, weather information, and lists ($p < .05$). The participants also responded to open-ended questions on the PSQ. While the participants' comments regarding the new palette were generally positive, they also had some common issues.

The weather presentation was a common theme among the PSQ open-ended responses. The participants stated that weather was more prominent with the new color standard. They also noted a preference for the single stipple pattern that we used to display weather with the new color standard. The participants reported that the single stipple pattern made it easier to distinguish between levels of weather intensity, and data blocks became more legible when they were superimposed on weather. However, some participants thought that the legibility of green (unowned) data blocks decreased with the new color standard when the data blocks overlapped with the “weather green” background color (i.e., weather severity levels 1 and 2).

Another common theme among the participants’ comments on the PSQ related to the color of the Air Traffic Proximity Alert (ATPA) cones. The ATPA cones change color as the distance between the lead and trailing aircraft changes. When ATPA is active and the system predicts that a trailing aircraft will lose minimum separation from a leading aircraft on approach or minimum separation has already been lost, the ATPA cone displays orange. However, the legacy color standard has been adapted over time such that the ATPA cone, while still technically orange, appears red. The participants noticed and commented on the distinct difference between the legacy color standard ATPA orange, and the new color standard ATPA orange that we implemented. In particular, the participants thought that the orange ATPA cones in the new color standard were less conspicuous than the “red” cones in the legacy color standard.

Table 11. Median (*Mdn*) rating and interquartile range (*IQR*) for PSQ items (Note: * statistically significant difference)

PSQ Item	**Legacy** ***Mdn (IQR)***	**New** ***Mdn (IQR)***
Please rate how the color palette affected your use of the Primary Control Monitor (PCM):		
1. Safety of Operations	4 (3)	5 (3)
2. Efficiency of Operations	4 (3)	4.5 (3)
3. Awareness of Potential Conflict	4 (3)	4.5 (3)
4. Ability to Coordinate with other sectors	4 (3)	4.5 (3)
5. Workload	4 (3)	4 (3)
Please rate the legibility of the following objects on the PCM during the scenario you just completed:		
*6. Target Symbols	6 (3)	7 (2)
7. Position Symbols	5.5 (3)	6.5 (2)
8. History Trails	5 (3)	7 (2)
9. Leader Lines	4 (3)	7 (2)
10. Vector Lines	4 (2)	6 (2)
11. Data Blocks: ACID	6 (3)	7 (2)
12. Data Blocks: Altitude	6 (3)	7 (2)
13. Data Blocks: Heading	7 (3)	7 (0)
14. Data Blocks: Speed	6 (3)	7 (2)
15. Data Blocks: Beacon Code	5 (3)	7 (3)
16. Data Blocks: Scratchpad	6 (3)	7 (2)
17. Data Blocks: Destination	6 (3)	7 (2)
18. Range Rings	6 (3)	6 (2.25)
19. Airports	6 (3)	7 (2)
20. Geographic Information	6 (3)	7 (1.5)
21. Minimum Vectoring Altitudes	7 (3)	7 (2)
22. Sector Boundaries	6 (3)	7 (2)
23. Sector Obstacles	6 (3)	7 (2)
24. Airways	4 (3)	7 (2.25)
*25. Weather Information	5 (3)	7 (1)
26. Air Traffic Proximity Alerts	6.5 (3)	7 (2)
27. Conflict Alerts	7 (3)	7 (2)
28. Minimum Safe Altitude Warnings	7 (3)	7 (2)

PSQ Item	Legacy *Mdn (IQR)*	New *Mdn (IQR)*
29. Toolbars	4 (3)	7 (3)
*30. Lists	4 (3)	7 (2.25)
31. J-Rings	6 (3)	6 (3)

The current study used a small sample size and some individual participant responses did not correspond with overall ratings. One participant thought that the aircraft position symbols were not as legible with the new color standard. However, the participants' PSQ ratings reflected a higher median rating (*Mdn* = 6.5) for position symbols when they used the new color standard compared to the legacy color standard (*Mdn* = 5.5). The participants also commented that some display elements and tools were too bright when they used the new color standard. We prevented the participants from adjusting the display brightness settings in a way that violated the 3:1 foreground-to-background luminance contrast ratio. Therefore, if a participant had increased the brightness of the background weather elements, the new color standard may have been prevented them from decreasing the brightness of foreground display elements and tools, depending on how bright the background weather elements were. To decrease the brightness of foreground display elements and tools, the participants also had to decrease the brightness of the background weather elements. The participants did not set the background weather elements to the minimum brightness settings, so either they required the background weather elements to be brighter than the minimum setting, or they did not fully understand the restrictions of the new color standard and the relationship between brightness of foreground and background elements. We find the latter explanation to be more likely as the weather depicted in the scenarios did not have a significant impact on air traffic management. The first time the participants interacted with the new color standard was during the data collection scenarios. While the participants understood the basic principle of foreground-to-background luminance-contrast ratios, the participants did not receive formal training on how to use and adjust the foreground and background-element brightness settings.

Although the participants reported that the colors were consistent between the legacy and new color standards, the 3:1 foreground-to-background luminance contrast restriction imposed by the new color standard required the participants to create new and unique display preference sets. The fact that some participants had issues finding the ideal display settings suggests that the participants may not have fully understood the interaction between foreground and background brightness settings. ATCS will need time at their facilities to create and try new display preference sets after implementation of the new color standard. All participant written comments and responses to open-ended items are contained in Appendix F.

3.5.2 Post experiment questionnaire

The PEQ used a 7-point Likert scale. For PEQ items 1 through 7, a rating of "1" represented "Extremely Inadequate" and a rating of "7" represented "Extremely Adequate", with the exception of Item 6 which used a scale where "1" represented "A Great Deal" and "7" represented "Not At All". PSQ items 8 through 13 were open-ended questions. Table 12 presents the median ratings and interquartile ranges for the PEQ items.

Table 12. Median (*Mdn*) rating and interquartile range (*IQR*) for the PEQ items

Item	***M (IQR)***
1. Rate the overall realism of the Air Traffic Control (ATC) simulation compared to actual ATC operations.	3.5 (2-6)
2. Rate the realism of the simulation hardware compared to actual equipment.	6 (4-7)
3. Rate the realism of the simulation software compared to actual functionality.	5 (3-6)
4. Rate the realism of the airspace compared to actual NAS airspace.	6.5 (6-7)
5. Rate the realism of the simulation traffic scenarios compared to actual NAS traffic.	6 (4-7)
6. To what extent did the WAK online workload rating technique interfere with your ATC duties?	6.5 (2-7)
7. How adequate was the airspace training?	6.5 (5-7)

The participants rated the overall realism of the simulation as moderate. They cited some differences from reality with aircraft performance, liveries, limited diversity of aircraft types, and simulation pilot idiosyncrasies. However, the participants rated the various aspects of the simulation as being higher in realism. They said the simulation hardware and software were realistic, but they were unable to perform some functions (e.g., depicting geographical areas). The participants thought that the airspace and air traffic scenarios were realistic, but again mentioned that some aircraft performance was unusual. The participants said that the WAK did not affect their ATC duties and was at most a "minor interference." The participants also agreed that the classroom and hands-on airspace training was adequate.

The participants responded in the open-ended questions that the new color standard did not affect their ability to control air traffic and the use of color between the legacy and new color standard was consistent (with exception of the weather and ATPA cones). When asked about the presence of any persistent color-related annoyances or distractions, the participants responded with the same issues as in the PSQ. In particular, they mentioned the orange ATPA cones that differed substantially in color from the "red" ATPA cones afforded by the legacy color standard and the perceived increase in overall brightness of the foreground display tools and elements.

The participants disagreed somewhat about adjustability of the PCM brightness. Some participants said that they were able to adjust the brightness of the PCM to a comfortable level, but others thought that the new color standard caused some PCM elements to be too bright. As previously discussed, the participants may have had some difficulty adjusting some display elements due to the 3:1 foreground-to-background luminance contrast ratio that we imposed in the new color standard condition. The brightness restrictions of the new color standard could cause some difficulty for ATCS in adjusting the brightness levels of some foreground display tools, at least initially. The new color standard may also force some ATCS to change the way they work to an extent. For example, some ATCS have developed novel methods for using the existing tools in a way would not be possible with the new color standard (e.g., dimming a display element to zero), if the FAA implemented the color standard in the same way we did. However, the FAA may decide to implement the color standard by categorizing and treating various display elements in a different way than the experiment. For example, when implementing the color standard, the FAA may determine that J-rings are not critical information and should not be considered as a foreground element subject to brightness restrictions. The participants also mentioned some artifacts related to our graphical implementation of the new color standard, but these issues did not directly affect our usability assessment of the new color standard.

When asked what aspects the FAA should consider when implementing a new color standard for ATC, the participants made a few suggestions. The participants said that ATCS should receive adequate training to understand the changes and limitations associated with the new color standard, and they should have time to develop and save new display preference sets. In line with common human factors practice, they also recommended that the FAA consider consistency across displays and platforms. The participants generally liked the weather colors included in the new color standard. However, some said they had difficulty reading unowned (green) data blocks when the data blocks superimposed on the background green weather color. Other suggestions provided by the participants for implementing the new color standard (e.g., color-coding of alarms and alerts; eliminating or reducing a 3:1 minimum foreground-to-background luminance contrast ratio) would violate existing human factors design standards and are not discussed in further detail. The complete set of participant written responses to open-ended items are contained in Appendix G.

4 Conclusion & recommendations

We implemented the legacy and new color standards on a TRACON PCM and asked experienced ATCS to control air traffic in a high-fidelity simulation using each standard. We did not detect any differences in objective measures of number of controller-to-pilot communications, airspace activity, or subjective workload ratings between the legacy and new color standard conditions. The participants rated their subjective workload as low in both conditions.

The participants rated the legibility of objects on the PCM as moderate to high for both color standards. However, the participants thought that the new color standard provided greater legibility for target symbols, weather information, and lists. The participants' PSQ responses and objective data suggest that neither the legacy nor new color standard impacted the safety or efficiency of operations, awareness of potential conflicts, ability to coordinate with other sectors, or workload. The participants did comment on some noticeable differences between the standards including the color of ATPA cones and restricted brightness adjustments. The participants made about twice as many display brightness adjustments when they used the new color standard vs. the legacy color standard, but the objective and subjective data supports the conclusion that the new color standard did not have adverse effects on the participants' ability to control air traffic.

The participants generally liked the weather presentation of the new color standard, but one participant commented that unowned data blocks could be difficult to see in when superimposed on the low-level weather background color. The additional background weather colors included in the new color standard required a novel design for presenting weather information. We based our presentation of weather information using the new color standard on rapid prototyping activities and SME feedback received during shakedown exercises for this study. We received positive feedback regarding the weather presentation with the new color standard.

The new color standard presented a color palette that was largely consistent with the legacy color palette. However, the restrictions on foreground-to-background luminance contrast imposed by the new color standard was novel. The dependency between foreground and background element brightness seemed intuitive to the design team but was not familiar to the participants. The participants had some difficulty adjusting overall display brightness, sometimes finding the display to be “too bright.” Some participants also discovered that they lost the ability to use some display strategies that relied on the ability to dim a display element completely.

In the current study, we applied the display standard to some elements (e.g., J-rings) that may or may not be subject to luminance contrast restrictions if the FAA implements the new color standard on existing systems. It is important to note that the means and methods we used to implement the new color standard in DESIREE differ significantly from operational implementation, especially in terms of the underlying software code. While the new color standard presents overall clear and logical information, implementing the color standard and restrictions on luminance contrast ratios to existing systems requires significant software development and testing. Once implemented, we recommend ATCS receive training on the new color standard regarding the use and limitations of PCM brightness adjustments and the dependencies between foreground and background elements. Additionally, ATCS will require some time to train on the new color standard at their facility so they can create and save new PCM preference sets for operational use.

Our implementation of the new color standard on a TRACON PCM showed no evidence of adverse effects on ATC efficiency or safety. Research on the design and presentation of weather information with the new color standard may improve the presentation of information on the TRACON PCM. We recommend further research on the design of weather information on the terminal PCM to determine the best method of weather presentation. The FAA must also consider the practical implications and challenges of implementing the new color standard on existing systems and software.

5 Acknowledgements

The Federal Aviation Administration Human-Systems Integration Branch (ANG-E5B) can accomplish studies such as this because of outstanding daily support from sponsors, ATCS SMEs, computer scientists, and support staff of the RDHFL. We would like to thank the FAA Human Factors Division (ANG-C1) and their Program Managers Karl Kaufmann and Dan Herschler for funding and overseeing this study. We also thank the FAA Air Traffic Organization Program Management Office Program Control and Integration Human Factors Lead, Ben Willems (AJM-131), for providing technical sponsorship and guidance. Many special and well-earned thanks to the dedicated and skillful team that made this study possible.

- Dr. Kenneth Allendoerfer – Manager – Human-Systems Integration Branch, ANG-E5B
- Albert Macias – RDHFL Team Lead – Concepts & Systems Integration Branch, ANG-E14
- Nigel Archibald –Manager – TGF, ANG-E16
- Daniel D’Ascenzo – DESIREE Software Engineer – Concepts & Systems Integration Branch, ANG-E14
- Laura Hamann - DESIREE Software Engineer- Concepts & Systems Integration Branch, ANG-E14
- Jennifer Librizzi – Secretary – Human-Systems Integration Branch, ANG-E5B
- Nelson Brown – Computer Scientist – Human-Systems Integration Branch, ANG-E5B
- Elena St. Amour – Engineering Research Psychologist – Human-Systems Integration Branch, ANG-E5B
- Adam Rhodes – ATCS, NATCA NextGen Representative
- Lee Moore – ATCS, Houston TRACON, NATCA Representative – Terminal CHI Team
- Eduard Antonius, Hareem Bohkari, Vincent Bryant, Eric Einhaus, Sam Isidoro, David Kazdov, Keith Maggio, Robert Mullin, Bryan Nunez, Alex Perez, Tyler Reich, Muhammad Sheikh, Laura Taneyhill – DESIREE Software Engineers – SAIC
- David L. Post –Color Science SME - Φ
- Ed Little – RDHFL Network Administrator - SAIC
- John Dilks – RDHFL Engineer – SAIC
- Fred Caraballo – RDHFL Network Administrator – SAIC
- Michael Perseo – Senior Systems Engineer - SAIC
- Josue Espinosa – RDHFL Engineer – ANG-E16
- Otto Smith – RDHFL Engineer - SAIC
- Alex Arrington – Computer Scientist - TGF, ANG-E16
- Karl Trout – Computer Scientist - TGF, SAIC
- Mai Truong – Computer Scientist – TGF, SAIC

- Michael Thomas – ATC SME, LST
- Eric Kirner – ATC SME, LST
- Ed Guguski – Contracting Officers Representative – Enterprise Testing and Capability Branch, ANG-E57

6 References

Chidester, T., Milburn, N., Lomangino, N., Baxter, N., Hughes, S., & Peterson, L. S. (2011). *Development, validation, and deployment of an occupational test of color vision for air traffic control specialists [DOT/FAA/AM-11/8]*. Washington, DC: Federal Aviation Administration.

Chidester, T., Milburn, N., Peterson, L. S., Gildea, K., Roberts, C., & Perry, D. (2013). *Development, validation, and deployment of a revised air traffic color vision test: Incorporationg advanced technologies and oceanic procedures and en route automation modernization systems [DOT/FAA/AM-13/18]*. Washington, DC: Federal Aviation Administration.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, J. (1992). Statistics: A power primer. *Psychological Bulletin, 112*, 155-159.

Federal Aviation Administration. (2007). *Baseline requirements for color use in air traffic control displays (FAA HF-STD-002)*. Washington, DC: Federal Aviation Administration.

Federal Aviation Administration. (2020). *Color use in Air Traffic Control system displays (FAA HF-STD-010A)*. Washington, CD: Federal Aviation Administration.

Geisser, S., & Greenhouse, S. W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics, 29*, 885-891.

Gildea, K., Milburn, N., & Post, D. L. (2018). *Development of a standard palette for color coding ATC displays (DOT/FAA/AM-18/22)*. Washington, DC: Federal Aviation Administration.

Gildea, K., Willems, B., Benincasa, S., Jack, D. G., & Post, D. L. (2020). *Evaluation of a new color palette for ATC displays (DOT/FAA/AM-20/08)*. Washington, DC: Federal Aviation Administration.

Hays, W. L. (1988). *Statistics* (4th ed.). New York: Harcourt Brace Jovanovich College Publishers.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measures designs have exact F distributions. *Journal of the American Statistical Association, 65*, 1582-1589.

IEC. (1999). *Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB (IEC 61966-2-1:1999).* Geneva, Switzerland: International Electrotechnical Commission.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Belmont, CA: Brooks-Cole.

Post, D. L., & Goode, W. E. (2020). Palette designer: A color-code design tool. *Displays, 61*, 1-6.

Truitt, T. R., Zingale, C. M., & Konkel, A. (2016). *UAS operational assessment visual compliance - Human-in-the-loop simulation to assess how UAS integration in class C airspace will affect air traffic control specialists [DOT/FAA/TC-16/11].* Washington, DC: Federal Aviation Administration.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80-83.

A Informed consent statement

Informed Consent to Participate in Research Study

Color Palette Operational Assessment

Principal Investigator (PI): Todd R. Truitt, Ph.D. – Federal Aviation Administration
Co-investigators: Brion Woroch, Ph.D., SAIC
Sponsors: Ben Willems - PMO Integrated Services & Analysis, Planning & Analysis Team (AJM-131)

Invitation to Participate in Research Study

Todd Truitt invites you to participate in a research study about the FAA Color Standard at the William J. Hughes Technical Center. The purpose of the study is to conduct a high-fidelity simulation to assess the usability of the new color palette standard for the terminal primary control display. The study will identify usability issues that occur when using the new color palette standard, propose remedies, inform future research, and support implementation decisions. Six Air Traffic Control Specialists will participate in the study. The study will follow the facility COVID protocol in place at the time of testing. All research personnel and participants must be compliant with agency requirements for access to the research site and the safety and health procedures in effect at the time of data collection. If the WJHTC is classified as "High" by the CDC COVID-19 Community Level criteria for Atlantic County, NJ, in-person data collection may be conducted only if the study has been designated "mission critical." If the facility classification changes during your visit, the research team will inform you and provide guidance on any changes in procedures.

Description of Participant Involvement

Each participant will be a Full Performance Level, Certified Professional Controller from a Level 10 or above Terminal Radar Approach Control (TRACON) facility. All participants must have normal color vision and normal, or corrected to normal, vision.

The participants will arrive at the William J. Hughes Technical Center in groups of two and will participate over 3 days. Each participant will complete air traffic control tasks at two different positions. The first day of the study will consist of a study briefing, equipment familiarization, and practice scenarios. Each scenario will last approximately 30 minutes. Data collection will begin on the second day and continue through the third day. We will run any make-up scenarios as needed on the third day and conduct a final debriefing. The participants will work from about 8:30 AM to about 4:30 PM every day with a lunch break and at least two rest breaks.

The participants will control TRACON traffic under two conditions – legacy color palette and new color palette. The participants will provide online ratings of subjective workload during each scenario. The participants will also have an "Easy Button" at each position that they can press to mark a point in the scenario for later review. After each scenario, the participants will complete a survey to evaluate the usability of the color palette. Subject Matter Experts and experimenters will observe and take notes during each scenario to further assess the color palettes. An automated data collection system will record system operations and generate a set of standard measures including safety, efficiency, and communications.

We will also collect audio-video data during the study so researchers can derive objective measures and reexamine any important events later if needed. All recordings will be stored in a secured location so that only members of the research team will have access to them. We will not publish or present any audio or video recordings.

Potential Benefits

I understand that I will be able to provide the researchers with valuable feedback and insight into the usability of the new color palette standard. My data will help the FAA establish the feasibility of implementing the new color palette standard for TRACON primary displays. I understand that the only benefit to me is that I will be able to provide the researchers with feedback and insight regarding the new color palette. My data will help the FAA to identify potential human factors issues with the color palette and help inform FAA standards for color palette implementation.

Risks and Discomforts

As a participant in this study, I understand that I will not be exposed to any intrusive measurement techniques. I understand that I will not be exposed to any foreseeable risks beyond what I usually experience in my every day job.

Participant's Rights

You will not lose any legal claims, rights, or remedies by signing this form and by your participation in this research study.

The local FAA Institutional Review Board has reviewed this research project under expedited review and found it to be acceptable, according to applicable state and federal regulations designed to protect the rights and welfare of subjects in research.

Confidentiality

The data collected and the recordings obtained in this study are stored only by code number, not by name. No names or identities will be released in any research reports, publications, or presentations resulting from this work. Electronic data will be maintained on secure FAA or FAA-contractor computers and websites that are accessible only by research team members. Any data collected on paper (e.g., questionnaires) will be secured in a locked file cabinet accessible only by research team members. The data from the study may be made available to other researchers for related studies. We will keep your participation in this research study confidential to the extent permitted by law.

Injury

In the event of any injury incurred while participating in this study, medical treatment will be provided by emergency responders, local hospitals, or clinics. Notify one of the researchers immediately if you need medical attention.

Voluntary Nature of Participation and Withdrawal

Your participation in this study is voluntary and it is your choice whether to participate or not. You may decline or withdraw your participation in the study at any time. The choice to decline or withdraw from the study will not cause any penalty or loss of any benefit to which you are entitled. During the study, the principal investigator or research team member will share any new information that develops that may affect your decision to continue to participate. The PI or research team may also terminate your participation in the study at any time if they determine this to be in your best interest.

Contact Information

If you have questions about the study, please ask them before signing this form. You can ask any questions that you have about this study at any time, or after your participation concludes.
For questions, concerns or complaints about this study, please contact the principal investigator, (Todd Truitt at 609-485-4351). If you feel that you have been treated unfairly, or you have questions regarding your rights as a research participant, you may contact the Local Institutional Review Board at 609-485-8629 or the FAA IRB at (405)954-2700.

Signature and Consent to be in the Research Study

I have been informed about the purpose, procedures, possible benefits, and risks of this research study. I have read (or someone has read to me) this form, and I have received a copy of it. I have had the opportunity to ask questions and to discuss the study with an investigator. My questions have been answered to my satisfaction. I have been told that I can ask other questions any time. I voluntarily agree to participate in this study. I am free to withdraw from this study at any time without penalty and without the need to justify my decision. The withdrawal will not in any way affect any benefits to which I am otherwise entitled. I agree to cooperate with the principal investigator and the research staff and to inform them immediately if I experience any unexpected or unusual symptoms.

Participant: By signing this consent form, you indicate that you are voluntarily choosing to take part in this research.

Printed Name of Participant

____________________________________ ________________

Signature of Participant Date

Investigator

I have fully explained this study to the subject to the best of my ability. As a representative of this study, I have explained the purpose, the procedures, the possible benefits, and risks that are involved in this research study. I have answered the subject's questions to his/her satisfaction before requesting the signature(s) above. I confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily. There are no blanks in this document. A copy of this form has been given to the subject.

____

Printed name of Principal Investigator

__________________________________ ________________ ____________

Signature of Principal Investigator Date Time

B Background questionnaire

Terminal Color Palette Operational Usability Assessment
Background Questionnaire

Please read the information below before completing this questionnaire!

This questionnaire collects general information about your experience as an Air Traffic Control Specialist. Researchers will only use this information to describe the participants in this study as a group. Researchers will not correlate or otherwise link information on this form to any data collected during the study. Your identity will remain anonymous.

1. How long have you worked as an **ATCS (include FAA developmental, CPC, and military experience)**?	_____ years _____ months

2. How many of the **past 12 months** have you actively controlled traffic in a TRACON facility?	_____ months

3. How long have you controlled traffic **using STARS**?	_____ years _____ months

C Post scenario questionnaire

Terminal Color Palette - Operational Usability Assessment
Post Scenario Questionnaire

Please rate how the color palette affected your use of the Primary Control Monitor (PCM).
During the scenario you just completed, how did the color palette affect the following?

1. **Safety of Operations** Very Negatively ① ② ③ ④ ⑤ ⑥ ⑦ Very Positively

Comments: ____
____
____
____

2. **Efficiency of Operations** Very Negatively ① ② ③ ④ ⑤ ⑥ ⑦ Very Positively

Comments: ____
____
____
____

3. **Awareness of potential conflicts** Very Negatively ① ② ③ ④ ⑤ ⑥ ⑦ Very Positively

Comments: ____
____
____
____

4. **Ability to Coordinate with other sectors** Very Negatively ① ② ③ ④ ⑤ ⑥ ⑦ Very Positively

Comments: ____
____
____
____

5. **Workload** Very Negatively ① ② ③ ④ ⑤ ⑥ ⑦ Very Positively

Comments: ____

____

____

____

Please rate the legibility of the following objects on the PCM during the scenario you just completed:				
6. **Target Symbols**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
7. **Position Symbols**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
8. **History Trails**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
9. **Leader Lines**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
10. **Vector Lines**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
11. **Data Blocks: ACID**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
12. **Data Blocks: Altitude**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
13. **Data Blocks: Heading**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
14. **Data Blocks: Speed**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
15. **Data Blocks: Beacon Code**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
16. **Data Blocks: Scratchpad**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
17. **Data Blocks: Destination**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
18. **Range Rings**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
19. **Airports**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
20. **Geographic Information**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
21. **Minimum Vectoring Altitudes**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A

Please rate the legibility of the following objects on the PCM during the scenario you just completed:				
22. **Sector Boundaries**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
23. **Sector Obstacles**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
24. **Airways**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
25. **Weather Information**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
26. **Air Traffic Proximity Alerts**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
27. **Conflict Alerts**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
28. **Minimum Safe Altitude Warnings**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
29. **Toolbars**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
30. **Lists**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A
31. **J-Rings**	Very Low	① ② ③ ④ ⑤ ⑥ ⑦	Very High	N/A

Please tell us if the color palette affected your use of the PCM.

32. **Did the color palette cause any problems?**

____

____

____

____

33. **Other comments**

____

____

____

____

D Post experiment questionnaire

Terminal Color Palette - Operational Usability Assessment
Post Experiment Questionnaire

Instructions: Please complete the following items based upon your overall experience in the study you just completed.

Simulation Realism and Research Equipment			
1. Rate the **overall realism of the Air Traffic Control (ATC) simulation** compared to actual ATC operations.	Extremely Unrealistic	① ② ③ ④ ⑤ ⑥ ⑦	Extremely Realistic

Comments: ____

____

____

____

2. Rate the **realism of the simulation hardware** compared to actual equipment.	Extremely Unrealistic	① ② ③ ④ ⑤ ⑥ ⑦	Extremely Realistic

Comments: ____

____

____

____

3. Rate the **realism of the simulation software** compared to actual functionality.	Extremely Unrealistic	① ② ③ ④ ⑤ ⑥ ⑦	Extremely Realistic

Comments: ____

____

____

____

4. Rate the **realism of the airspace** compared to actual NAS airspace.	Extremely Unrealistic	① ② ③ ④ ⑤ ⑥ ⑦	Extremely Realistic

Comments: ____

____

____

____

5. Rate the **realism of the simulation traffic scenarios** compared to actual NAS traffic.	Extremely Unrealistic	① ② ③ ④ ⑤ ⑥ ⑦	Extremely Realistic

Comments: ____

____

____

____

6. To what extent did the **WAK online workload rating technique interfere** with your ATC duties?	A Great Deal	① ② ③ ④ ⑤ ⑥ ⑦	Not At All

Comments: ____

____

____

____

7. How adequate was the **airspace training**?	Extremely Inadequate	① ② ③ ④ ⑤ ⑥ ⑦	Extremely Adequate

Comments: ____

____

____

____

Color Palette

8. **Did the new color palette affect your ability to control air traffic? If so, how?**

____

____

____

____

____

____

____

____

____

____

____

____

____

____

9. **Was the use of color consistent? If not, what discrepancies did you notice?**

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

10. **Were there any persistent color-related annoyances or distractions during the study?**

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

11. **Was the Primary Control Monitor adjustable to a comfortable brightness?**

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

12. **What aspects should the FAA consider when implementing a new color palette for ATC?**

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

13. **Do you have any additional comments regarding the study?**

E WAK instructions

(The full set of instructions will be read at the beginning of each test day). An abbreviated set of instructions will be read prior to each experimental run. The abbreviated instructions will omit the first paragraph below.)

One purpose of this research is to obtain an accurate evaluation of controller workload. By workload, we mean all the physical and mental effort that you must exert to do your job. This includes maintaining the “picture,” planning, coordinating, decision making, communicating, and whatever else is required to maintain a safe and expeditious traffic flow. Workload is your perception of how hard you must work to perform all of the tasks necessary to meet these demands, not necessarily a measure of how much traffic you are working. Workload levels fluctuate. All controllers, no matter how proficient, will experience all levels of workload at one time or another. It does not detract from a controller’s professionalism when he indicates that he is working very hard at certain times or that he is hardly working at other times.

Every 4 minutes the WAK device, located at your position, will emit a brief tone and the 10 buttons will illuminate. The buttons will remain lit for 20 seconds. Please tell us what your workload is at that moment by pushing one of the buttons numbered from 1 to 10.

At the low end of the scale (1 or 2), your workload is low - you can accomplish everything easily. As the numbers increase, your workload is getting higher. The numbers 3, 4, and 5 represent increasing levels of moderate workload where the chance of error is still low but steadily increasing. The numbers 6, 7, and 8 reflect relatively high workload where there is some chance of making errors. At the high end of the scale are the numbers 9 and 10, which represent a very high workload, where it is likely that you will have to leave some tasks unfinished. Feel free to use the entire rating scale and tell us honestly how hard you are working at the instant that you are prompted. Do not sacrifice the safe and expeditious flow of traffic in order to respond to the WAK device.

F Participant comments – Post scenario questionnaire

Please rate how the color palette affected your use of the Primary Control Monitor (PCM).

Item 1. Safety of Operations

Legacy Palette

New Palette

- ATPA cones are not as prominent. Green tags don't show as well in WX areas
- I don't care for the orange cones over the red
- Alert cones for ATPA. I like them red instead of orange
- Don't like the alert cone colors
- Colors didn't impact safety. I didn't like the orange ATPA cone as much as red, it didn't "pop" as much.
- The WX stood out more
- Just noticed the WX completely covers the MVA and airspace maps

Item 2. Efficiency of Operations

Legacy Palette

- [Improved legibility on new palette

New Palette

- Orange ATPA cones aren't as dominant or attention grabbing
- Interpreting precipitation levels

Item 3. Awareness of potential conflicts

Legacy Palette

- Red cones are better for ATPA

New Palette

- ATPA orange wasn't as dominant as red
- ATPA cone being orange

Item 4. Ability to Coordinate with other sectors

Legacy Palette

New Palette

- The tower did not answer me

Item 5. Workload

- No comments

Item 32. Did the color palette cause any problems?

Legacy Palette

- No problems. The lighter green from alternate scheme is preferred except when on a green tag in light precipitation.
- The standard WX stipples inhibit the visibility of planes/details underneath
- Old style WX compared to the newer colors makes it more difficult to distinguish between WX lvl 3+
- The standard WX stipples inhibit the visibility of planes/details underneath

New Palette

- No problems. The weather was definitely brighter and more prominent. The ATPA was not as prominent though.
- Green tags less visible in light precip.
- No problems, Alert cones aren't as attention drawing in these colors
- No problems. ATPA cones don't have same "draw" in orange as red to me. We've been conditioned RED is an imminent situation "CA", "LA', etc. Orange doesn't seem as urgent
- I liked the new colors for WX. It makes heavy (3,4) and extreme (5,6) much more easily distinguishable. The J-rings and P bats were too bright and I couldn't find a way to dim them. The List brightness wouldn't go below 25. I would turn that down more.
- The tools menu is way too bright turned all the way down. P-cones, J-rings, etc. are too bright.
- The J-rings, P-cones, etc. are too bright on the lowest setting. Rand

rings even when off are still displayed throughout the WX. Compass even when off is still displayed through the WX.

- Some things couldn't be dimmed / removed completely such as range rings. Even on 0 they were still visible. Compass rose, when set to off, was still visible through the precipitation
- No, but I would like to dim colors overall more
- The brightness of range rings and J-rings, P-cones seem to be brighter than the same pref. sets and doesn't carry over correctly from the legacy. Position symbols seem to be more difficult to read and differentiate.
-
- The range rings go from dim @ 2 or 3 then one click up the brightness is high. There is not a gradual brightness increase as you add brightness. The tools and associated functions are too bright at minimum settings
- No, but even at the lowest brightness it's too bright for the WX, lists, DCB, etc.
- I liked the new colors for WX. It makes heavy (3,4) and extreme (5,6) much more easily distinguishable. The J-rings and P bats were too bright and I couldn't find a way to dim them. The List brightness

- wouldn't go below 25. I would turn that down more.
- None except WX overwriting map which is a simulation issue

Item 33. Other comments

Legacy Palette

- Having seen the alternate weather option, I do prefer the alternate. "Double stiples" are harder to distinguish than stiples over a different color
- Noticed more muted appearance on old color palette.
- ATPA seems to go in and out reading the correct aircraft

New Palette

- I'm not sure if the green is the same, but it looked paler. If it was different, I liked it.
- One callsign had too many digits (at beginning of scenario)
- The green is more subtle and is preferred. The blue seems more pale and I prefer the legacy for definition
- N1759CP has too many letters for a callsign
- The colors all seem OK to me
- I had issues with the DCB and the ability to make entries vis my slewball.
- Colors remained consistent with legacy conventions
- Seems more crisp
- Brightness values are much brighter than the same setting as legacy

G Participant comments – Post experiment questionnaire

1. Rate the overall realism of the Air Traffic Control (ATC) simulation compared to actual ATC operations.

- Some of the speeds were much slower than expected (PA28 should not catch BE36). Some a/c made turns out after cleared for approaches.
- It was realistic except for dome of the speeds inside the marker. Sometimes a PA28 would be overtaken by a BE36.
- Wasn't sure about 1200 codes and how they fly reference airspace surrounding OAK,HWD,SFO, I had to refrain from telling pilots to "fly heading #, intercept localizer" instead I would just issue a heading because if I said "intercept" the RPO command would negate my heading.
- As far as simulators go it looked realistic. It would have been nice to have different scenarios to keep interest in traffic
- Was overall good. The only party would be pilot error in some command entries, which is unavoidable. The pilots actually did a very good job overall keeping up.
- Aircraft types should be updated to current livery. Would be beneficial to have greater diversity of a/c types. Would be nice to have VFR pickups be added for realism.

2. Rate the realism of the simulation hardware compared to actual equipment.

- The equipment seemed to operate for the most part identical to STARS.
- I thought the equipment felt normal.
- Couldn't do some inputs such as automated point outs with adjacent sector.
- The hardware was very close to what is used in the field.
- Most of STARS commands are available.

3. Rate the realism of the simulation software compared to actual functionality.

- Everything was virtually identical to STARS for the functions I attempted.
- The
- If a plane is vectored at, or inside of the FAF, it will make a turn further out on the final. Some props don't have to configure for landing they shouldn't be expecting a "long" final.
- There was some functionality that was missing that would have been nice to test with the new color palette, geo areas specifically
- The software was functional enough to effectively complete the study.

4. Rate the realism of the airspace compared to actual NAS airspace.

- Not familiar with NCT airspace, but it seemed realistic.
- Airspace felt like a normal airspace.
- I'm assuming there's extensive Bravo airspace surrounding the airports in the scenario so at times it seemed there were conflicts with 1200 code traffic. That shouldn't be affecting planes on the rwy 30 final.
- Actual NCT traffic flow is uncertain to me but arrivals and overflights appeared to follow metro airspace.

5. Rate the realism of the simulation traffic scenarios compared to actual NAS traffic.

- I would assume this is close to OAK traffic. Performance characteristics of certain a/c seemed a little off, but generally good.
- There was WX displayed in all the scenarios which was good to play with the colors and brightness, but since the automated planes on victor airways were flying through it I didn't take the WX seriously.
- Not actually having experience with this sector in real life, I'm not sure as to how much there normally is. Flight speeds at some a/c seemed out of the ordinary,

6. To what extent did the WAK online workload rating technique interfere with your ATC duties?

- During some of the busiest moments I may not have heard the alert, but think I responded most of the time.
- It was not a factor.
- Sometimes I would be doing key pack work and have to stop to hit the button then resume working. It was a minor interference.

7. How adequate was the airspace training?

- We got a broad overview in class. The two practice problems were enough to understand enough how to run the traffic.
- Training was adequate,
- There was plenty of info provided but no way to retain it all immediately. Once I memorized the frequencies it was easier to focus on the scenarios so I wasn't constantly looking away or up for a frequency.
- I feel the airspace training was almost too much information for the simulations we were running.

8. **Did the new color palette affect your ability to control air traffic? If so, how?**

- It didn’t affect my ability to work. I did like the softer green tag, but thought they somewhat blended in with level 1 WX. I did not like orange alert cones, as they weren’t as attention grabbing as red. I thought the WX was easier to distinguish levels than looking for wide or closely space stipples.
- No it did not. I wasn’t the biggest fan of the orange bat instead of the red cone. This did not affect my ability to control.
- The new colors are nice, but I need to be able to dim everything much lower and can just turn a knob if I need it brighter for a moment. The new WX colors made it significantly easier to distinguish between various levels of weather. The new colors didn’t hurt, WX

was easier to see, planes near WX were able to see, but ultimately everything on the lowest setting is too bright.

- It did make the position symbols M & N more difficult to read. Also, the pref set brightness of certain items would carry over in number from legacy but would be much brighter. This would lead to adjustments needing to be made while working traffic. The orange color for ATPA was distracting, but I feel this is because it was a new color.
- No
- No significant change to controlling traffic. Critical data was legible and consistent with established conventions.

9. **Was the use of color consistent? If not, what discrepancies did you notice?**

- They seemed consistent when using new vs legacy problems.
- Yes, I felt like it was.
- It was consistent.
- I did not notice any color discrepancies during testing.
- Overall, it felt very consistent
- Yes, with the obvious exception of weather.

10. **Were there any persistent color-related annoyances or distractions during the study?**

- No, only thing I didn't like was the orange bats for ATPA.
- Just the orange bats. Not a big deal, just my preference.
- Other than being bright, no.
- Orange in the ATPA was distracting.
- The only one I noticed seems to be lab related where the weather would completely obscure the MVA and airspace maps.

11. **Was the Primary Control Monitor adjustable to a comfortable brightness?**

- Yes
- Yes, felt like a normal radar room.
- No
- It wasn't. I feel the functionality related to J-rings, P-cones, and other "tools" related items were too bright at the lowest setting. Also the inability to see geo-restriction areas, this is an open issue. Range rings when off were still visible through the WX displayed, as well as compass numbers on the edge of the scope.
- It was. I did notice that the WX for the new color palette stayed at a fairly high intensity for the lowest allowable brightness setting.
- Yes, however new palette intensity seemed brighter which required me to lower brightness settings. Controllers who set brightness fairly low may need to adjust.

12. **What aspects should the FAA consider when implementing a new color palette for ATC?**

- Keep red for alert items (CA, LA, ATPA alert). Make sure green tags don't blend in w/ WX. Make sure WX different colors rather than wide vs tight stipples (like the new palette).
- I felt like most of the new colors used. I like the WX better in simulation than the real world.
- Be able to dim to almost invisible and then we can increase the brightness momentarily if necessary.
- I think they should allow training as well as time to convert over pref sets so the adjustments aren't having to be made while working traffic.
- Just making sure the workforce is aware of the change before implementation.
- Consistency across platforms and equipment of the same platforms.

13. **Do you have any additional comments regarding the study?**

- Speed on the final need tweaked. Some planes turned around on a visual approach after being cleared. This is not like real life. Turn rates after clearance could use a tweak. They turned slower than normal.
- It became monotonous running the same scenario so many times, but on the other hand it gave more time to mess w/ settings since we started to memorize the scenarios.
- I really liked the WX display. The study was very thorough.