DOT/FAA/TC-19/36

Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

Traffic Flow Management (TFM) Tools: Guidance for Use, Integration, and Training: Part-Task Experiment 2

Brion Woroch, Ph.D., SAIC
Carolina M. Zingale, Ph.D., FAA Human Factors Branch, ANG-E25

September 2019

Technical Report

U.S. Department of Transportation

Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/TC-19/36		

4. Title and Subtitle	5. Report Date
Traffic Flow Management (TFM) Tools: Guidance for Use, Integration, and Training: Part Task Experiment 2	2019
	6. Performing Organization Code ANG-E25
7. Author(s) Brion Woroch, Ph,D., SAIC Carolina M. Zingale, Ph.D., FAA Human Factors Branch	8. Performing Organization Report No. DOT/FAA/TC-19/36
9. Performing Organization Name and Address Federal Aviation Administration Human Factors Branch 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 Research and Engineering Group 800 Independence Avenue, S.W. Washington, DC 20591	13. Type of Report and Period Covered Technical Note/Technical Report 14. Sponsoring Agency Code ANG-C1

15. Supplementary Notes

16. Abstract

Objective: We conducted the Phase II part-task experiment to evaluate the effects of different training methods on the effectiveness of Decision Support Tools (DSTs). The Phase I part-task experiment involved a training method that utilized both a slideshow and an interactive learning component. In Phase II, the participants only viewed the information in the training slideshow and did not interact with the information in the same way as the participants in Phase I. We evaluated whether interactive training resulted in better task performance than a more passive form of training. **Method**: Eight volunteers from the FAA William J. Hughes Technical Center with no experience with Traffic Flow Management tools and procedures served as participants. We collected data from this group of participants using the same methods we used in Phase I. We compared the performance of participants between Phase I and Phase II on an aircraft rerouting task and two secondary tasks. **Results:** We did not find evidence that the interactive training method improved task performance. However, we did find that interactive training led to increased performance on the initial practice scenario of the task. **Conclusion:** This benefit may have implications for real-world training. If the first impression of a new tool is positive, it may lead to increased adoption and future use. However, if the first impression of a new tool is negative, or if the tool is difficult to learn, users may abandon it or use it less often. **Applications**: We recommend that training materials, 1) allow the user to set the pace for learning, and 2) provide the user specific feedback about performance and demonstrate the potential benefits of new tools and procedures.

17. Key Words	18. Distribution Statement
Decision Support Tools Traffic Flow Management Air Traffic Control	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 Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	49	

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

THIS PAGE IS BLANK INTENTIONALLY.

Table of Contents

List of Illustrations

Figures **Page**

Tables

Acknowledgments

We thank Nelson Brown (ANG-E25), computer scientist, and Kevin Tracy, Brett Williams, and Louis Rivera of Hi-Tec Systems, for developing the software and data collection system used in both of the experiments conducted in support of this project. We thank Tanya Yuditsky (ANG-E25) for her feedback and guidance on Traffic Flow Management tools and their use. We thank the participants from various organizations throughout the FAA William J. Hughes Technical Center for contributing their time and effort in support of this project.

THIS PAGE IS BLANK INTENTIONALLY.

Executive Summary

We conducted the Phase II part-task experiment to evaluate the effects of different training methods on Decision Support Tool (DST) effectiveness. The Phase I part-task experiment focused on four factors in DST use; situation-specific training, DST reliability, the number of recommendations made by the DST, and overall task workload. That experiment involved a training method that utilized both a slideshow and an interactive learning component (Woroch, Zingale, & Masalonis, 2017). The training that we developed for Phase I involved more interactivity than many of the Traffic Flow Management (TFM) web-based and computer-based instruction (CBI) training materials we later reviewed (Zingale, Masalonis, Woroch, & Yuditsky, 2017).

The Phase II part-task experiment added a new comparison group to the existing data from Phase I. We created a new training slideshow to be more similar to the existing TFM CBI training materials. The participants in Phase II only viewed the information in the training slideshow and did not interact with the information in the same way as the participants in Phase I. To keep the comparison focused on interactivity versus no interactivity, the Phase II slideshow training contained identical information to the Phase I training slideshow but included additional screenshots that replaced the interactive component. In this document, we refer to the training approach used in Phase I as "CBI-Do" and the training approach used in Phase II as "CBI-View."

This experiment tested whether interactive training resulted in better task performance than a more passive form of training. We did not find evidence that the interactive training method improved task performance. However, we did find that interactive training led to increased performance on the initial practice scenario of the task. This benefit may have implications for real-world training. If the first impression of a new tool is positive and its perceived usefulness is high, that may lead to increased adoption and future use. However, if the first impression of a new tool is negative, or it the tool is difficult to learn, users may abandon it or use it less often. Future research on this topic would be informative. Our non-interactive training also implemented two features that we suggest would be useful additions to current operational CBIs. We recommend that training materials:

- Allow the user to set the pace for learning instead of requiring that the user proceed at the pace imposed by the automation.
- Provide the user specific feedback about performance and demonstrate the potential benefits of new tools and procedures.

THIS PAGE IS BLANK INTENTIONALLY.

1. INTRODUCTION

This document summarizes the results of the second part-task study conducted to investigate Decision Support Tools (DST) use. The decisions made by Traffic Flow Management (TFM) personnel are complex and involve multiple factors that affect cognitive workload. DSTs, including the specific types that are the focus of this research—recommendation tools that provide users with suggested solutions to problems, and "what-if" modeling tools that allow users to examine the likely outcomes of different potential actions and scenarios—are intended to reduce cognitive workload. However, various factors, such as user training, are important considerations in implementing the tools to ensure that they bring about the expected benefits.

The first part-task experiment (Phase I) focused on four factors in DST use; situation-specific training, DST reliability, the number of recommendations made by the DST, and overall task workload. All factors had an effect on both objective and subjective measures of task performance. Several of the factors interacted in meaningful ways that illustrate the complex nature of DST use. DST reliability, training, and task workload played important roles in performance.

The second part-task experiment (Phase II) is an addendum to the first part-task experiment. In Phase II, we investigated the impact of a different training method on DST use in the TFM environment using the same methods we used in Phase I and that are described in Woroch, Zingale, & Masalonis (2017).

The Phase I study used a part-task design to assess DST use by non-TFM personnel. We designed a task that focused on several key aspects of the types of tasks performed by TFM personnel but that novices could learn quickly. The primary task we designed was inspired by the Integrated Departure Route Planning tool (DeLaura et al., 2012; Davison Reynolds & DeLaura, 2011) and involved rerouting aircraft around severe weather. The DST we developed for our experiments recommended one or more routes as a "high-scoring" route. The participants scored points based upon the choices they made for the reroutes. We included a training manipulation in Phase I in which we compared two groups of participants; one that received an enhanced situation-specific training (SST) module that explained the circumstances under which recommendations would be more or less useful and one that did not. We found that the SST affected DST use by allowing participants to better utilize the automation, particularly in high workload situations.

In order to teach our novice participants how to use the DST software, we had to develop a training and practice protocol. The experimental task is complex, so we wanted our training to be as concise and informative as possible. We evaluated several of the current computer-based instruction (CBI) and web-based training materials and summarized our findings in the TFM Decision Support Tools Training Tools Assessment document (Zingale, Masalonis, Woroch, & Yuditsky, 2017). We observed that many of the CBI training modules had little interactive learning and primarily focused on informative slideshows and descriptions. For the current study, we named this approach CBI-View to emphasize that most of the learning is done by viewing the information.

We identified a need for users to have the opportunity to use and test different functions at their own pace. Means et al. (1998) reached a similar conclusion in an evaluation of ATC training systems for En Route controllers conducted by the FAA almost 30 years ago. They suggested that academic learning of the materials should be integrated with active simulation and on-the-job training. To maximize learning and retention, the information should be put into use during practice.

For our Phase I part-task experiment, we developed an interactive training protocol that included a slideshow, based upon current CBI materials, interspersed with hands-on practice with the experiment software using demonstration scenarios. This allowed the participants to both learn the task quickly and apply that knowledge at their own pace. We named this a CBI-Do approach, to emphasize that the learning is from both CBI and hands-on experience. We feel that the training we developed was very effective in teaching how to perform the task and use the DST. However, we do not have empirical evidence that this training method would have been any better than a method similar to many current CBI training approaches.

We hypothesize that CBI-Do training improves performance in comparison to CBI-View training. We tested this hypothesis by adding an additional experimental group to the previous Phase I study. We already collected data from 8 participants with a CBI-Do approach as part of the Phase I part-task experiment (Woroch, Zingale, and Masalonis (2017). In the current experiment, we modified the training materials to create a CBI-View training procedure and collected data from an additional 8 participants using the existing task. We compared the performance of our new "CBI-view" group with that of the existing CBI-Do group. This allowed us to compare a CBI-based training approach that mimics the way training is often conducted today, to a training approach that allows users more hands-on practice.

2. METHOD

The current experiment used the same experimental protocol used by Woroch, Zingale, and Masalonis (2017) in Phase I. We provide the same protocol description in this document to allow this report to serve as a stand-alone document. We leveraged the existing experimental paradigm to evaluate an additional experimental training condition with a new group of eight novice participants. We designed the original part-task experiment to examine the use of Decision Support Tools (DSTs) in a TFM-inspired task. We designed a primary task that involved rerouting aircraft around severe weather. The participants also performed two secondary tasks to increase their workload. The secondary tasks were also inspired by TFM tasks. The entire experiment, including training, took each participant approximately 2.5 hours to complete. The current Phase II experiment involved a new group of participants to examine how different training procedures influenced task performance. The Phase II participants performed the identical part-task experiment as the participants in Phase I, but with a different training protocol.

2.1 Participants

Eight volunteers (1 Female) from the FAA William J. Hughes Technical Center (WJHTC), with no experience with Traffic Flow Management, served as participants in this study. The participants were employees at the Technical Center in roles (e.g., computer scientists, research psychologists) unrelated to the research and development of TFM tools. The participants ranged in age from 24-53 (M = 38.25, SD = 12.15). We attempted to match the participants in Phase II with those in Phase I as closely as possible. As a result, we excluded one potential participant's data from the analysis because his age was considerably outside the range of the other participants. All participants read and signed an Informed Consent Statement (Appendix A) that summarized their rights and responsibilities before participating.

2.2 Materials

We used the same materials and the same protocol for the primary rerouting task and the secondary tasks in the Phase II experiment as we had in Phase I (Woroch, Zingale, and Masalonis, 2017). We repeat that information in this document in the following sections. However, we also included an additional survey in Phase II regarding software usability and the effectiveness of the training provided as noted.

2.2.1 Demographics Questionnaire

The participants completed the demographics questionnaire (Appendix B) before beginning the experiment. It included questions about age, gender, TFM or Air Traffic Control (ATC) experience. We also confirmed that participants had not received information from prior participants that might give them a performance advantage.

2.2.2 Complacency Rating Scale

Before the participant performed the experimental tasks, we administered a survey to assess individual differences in attitudes toward automation and susceptibility to overreliance on automation. We used the Complacency Potential Rating Scale (CPRS) (Singh, Molloy, & Parasuraman, 1993), but we adjusted two items due to that scale's outdated technology references (e.g., an item about VCRs). We made edits to questions #7 and #20 to update the items pertaining to television and medical devices. The version used in this study is Appendix C.

2.2.3 Questionnaires

As in Phase I, we included questionnaires during the experiment; after each reroute, after each scenario, and at the end of all scenarios. We provide screen shots of those surveys in Appendix D.

For the current experiment, we included an additional survey that we administered at the end of all the experimental scenarios. This survey consisted of questions related to the software's interface and usability as well as the effectiveness of the training materials. We provide this survey in Appendix D.

2.2.4 Primary Rerouting Task

The primary task for the participants was a departure-rerouting task we designed based on the Integrated Departure Route Planning tool (DeLaura et al., 2012; Davison Reynolds & DeLaura, 2011). The task consisted of ten experimental scenarios, each requiring the participant to reroute several aircraft around a weather event from one airport to another in fictional airspace. We based the task on tools a Traffic Manager might use. However, we greatly simplified both the number of parameters used and the task itself so a participant with no TFM experience could learn and practice the task and complete the ten experimental scenarios in 2-2.5 hours.

We provide a screenshot of the rerouting task in Figure 1. It contains a map of all the routes and a list of aircraft to be routed through that airspace. During each scenario, five aircraft had their originally filed route blocked by severe weather (e.g., Route E) and had to be rerouted to a different route. The airports, route names, waypoint labels, and direction of flight varied between scenarios but stayed the same for all five flights throughout a single scenario. Table 1 shows all of the parameters that varied between scenarios. Every scenario began at 1900 hours and each flight was scheduled to depart 15 minutes after the previous flight (1900, 1915, 1930, 1945, and 2000).

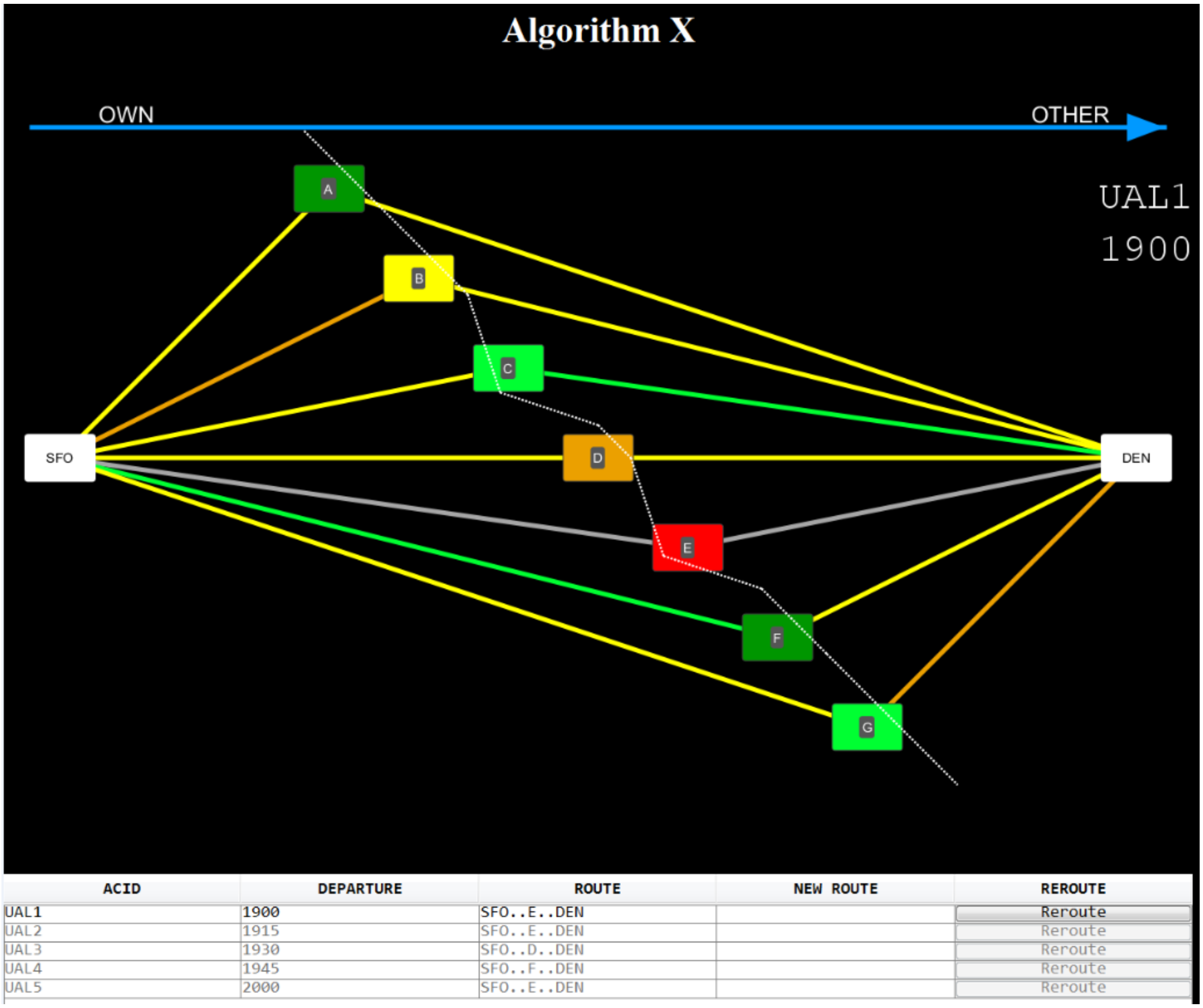
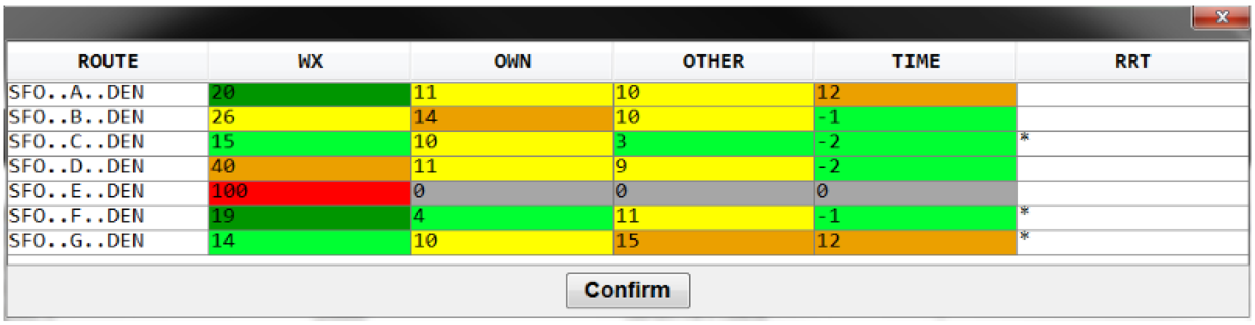


Figure 1. A screenshot of the primary aircraft-rerouting task.

Table 1. Scenario Variables

Airports	Traffic Direction	Aircraft ID (ACID)	Waypoint Labels	Number of Scenarios
SFO & DEN	Eastbound	UAL 1-5	A through G	5 (5 flights in each)
JFK & ORD	Westbound	JBU 1-5	H through N	5 (5 flights in each)

Each scenario featured seven routes from the departure airport to the destination airport. Each route passed through a single waypoint. At the start of each flight, the originally filed route was closed due to severe weather. For example, in Figure 1, UAL1 was originally filed as SFO..E..DEN. The participant had to choose an alternate route for the flight from among one of the six remaining routes. Each route was assigned a score. The score was based on four parameters that varied for each flight: weather, flight delay time, congestion level in the participant's center, and congestion level in the neighboring center. Each parameter was assigned a numeric value that was weighted based on their relative importance, as determined by the experimenters, and combined to yield a final score for each route. Figure 2 shows the route table that displayed the parameters for each route at the departure time of that aircraft. These numbers combined to yield a score out of 1000 possible points for each reroute. Additionally, a color scale indicated the relative severity (i.e., greatest negative impact to route) of that parameter. The colors used in this task; from lowest to highest, were: light green, dark green, yellow, and orange (Table 2). Red and gray indicated a closed route.



ROUTE	WX	OWN	OTHER	TIME	RRT
SFO..A..DEN	20	11	10	12	
SFO..B..DEN	26	14	10	-1	
SFO..C..DEN	15	10	3	-2	*
SFO..D..DEN	40	11	9	-2	
SFO..E..DEN	100	0	0	0	
SFO..F..DEN	19	4	11	-1	*
SFO..G..DEN	14	10	15	12	*

Confirm

Figure 2. The route table displaying scoring parameters.

Table 2. Route Parameter Values

	Weather	Own Center	Other Center	Delay Time
Gray	0	0	0	0
Light Green	1 to 15	1 to 4	1 to 4	-10 to -1
Dark Green	16 to 20	5 to 7	5 to 7	1 to 5
Yellow	21 to 35	8 to 11	8 to 11	6 to 10
Orange	36 to 50	12 to 15	12 to 15	11 to 20
Red	>50	N/A	N/A	N/A

<u>Weather:</u>

The originally filed route for each aircraft was blocked due to severe weather. This was represented visually on the route map (see Figure 1) with a red box around the waypoint label. This was represented in the route table (see Figure 2) with a red cell and "100" parameter value. The weather parameter for the other available routes varied from 1-50, with a higher number indicating more severe weather. The weather condition predicted for each route was indicated by the color of the waypoint on the map and the color of the cells in the route table. We instructed participants that all colors except red represented weather that was safe to fly in and that the lower the weather value the higher the score for that route.

Own & Other Center Congestion:

The airspace on the route map (see Figure 1) was separated into two halves, the airspace under the control of the participants **own** center and the airspace under the control of the **other** center. In scenarios in which traffic was eastbound from SFO to DEN, the airspace containing SFO and the routes to the left of the waypoint were **OWN** and the rest of the airspace was **OTHER**. In scenarios in which traffic was westbound from JFK to ORD, the airspace containing JFK and the routes to the right of the waypoints were **OWN** and the rest was of the airspace was **OTHER**. The own and other labels were shown at the top of the screen throughout the scenario.

The aircraft congestion level along the routes was indicated by the color of the route line and the color of the corresponding cells in the route table. We told the participants that the number in the route table indicated the number of aircraft forecasted along the route. Rerouting onto a lower number (less congested) route yielded more points.

Delay Time:

The delay time parameter was the number of minutes a flight would be delayed by choosing that route. This value varied from *-10* to *20*. A negative number meant a route was faster than the originally filed route and was a saving time. This parameter was not represented visually on the route map, it was only available in the route table. A faster flight time (less delay) yielded a higher score.

Score Calculation:

We developed the following formula to combine the individual parameters into a score for that route: ((70-Weather)*8.1)+((20-Own Center)*14)+((20-Other Center)*9)+((20-Time)*7.4). The factors were weighted to represent the relative importance of each parameter, as determined by the researchers based on our knowledge of TFM and discussions with TFM subject matter experts (SMEs) to reflect the relative importance of these factors in the real world (Woroch, Zingale, and Masalonis, 2017). We identified weather as the most important factor. A low value on the weather parameter contributed more to the score on that route than any other parameter. The second most important factor was the congestion level in the participant's **own** airspace. Congestion levels in the **other** airspace and flight delay contributed about equally to the score and were weighted the lowest.

We instructed the participants that the overall the goal of the task was to choose the highest scoring route to obtain the highest possible score in each scenario. We trained the participants on the parameters and how the score was calculated. After each scenario, participants received feedback regarding their choices and score.

2.2.4.1 Task Sequence and "What-iffing" Capabilities

The participants had four minutes to reroute the five aircraft in each scenario. The route table (see Figure 2) started as an empty table with no information except the indication that the originally filed route closed due to severe weather. The participants could gather information about other routes by selecting them with the mouse, a procedure we referred to as "what-iffing" because it allowed the participants to see the parameters associated with that route option. The alternate routes indicated the colors of the route lines and waypoints on the route map (see Figure 1). Choosing to "what-if" a route options was followed by a 3-second delay before the information appeared in the route table. This delay was added to mimic potential computer-processing time required by some tools to model information. The delay was also added to discourage participants from "what-iffing" all routes for all flights in every scenario. Because of the time pressure to complete the scenario within the allotted 4 minutes, the delay introduced a cost to acquiring more information and we hoped to encourage a more selective use of this feature.

All five aircraft had to be rerouted within four minutes, or the participants received a 500-point score penalty and an additional 10-point penalty for every ten seconds over 4 minutes. We did this to encourage the participants to finish the scenario within four minutes. However, we still wanted to gather any rerouting choices for flights rerouted after the four minutes expired, so the scenario continued to run until the participants completed all of the reroutes.

2.2.4.2 Route Recommendation Tool

Most of the scenarios (8 of 10) had routes suggested by the Route Recommendation Tool (RRT). A suggested route had the parameter information for that route automatically displayed when the rerouting for that flight began. Additionally, an asterisk was placed, in that route's row, in the RRT column of the routing table to indicate that it was a "recommended" route. Four of the ten scenarios had 1 route automatically recommended for each flight; four scenarios had 3 routes recommended for each flight. The two remaining scenarios had no automation suggestions. Figure 2 shows a route table with three routes recommended by the RRT and all other routes "what-iffed."

The RRT varied in the quality of its recommendations; sometimes it indicated the highest scoring route available and other times it did not. Thus, the RRT varied whether it would reliably recommend a good route. We systematically assigned which route was recommended, allowing the creation of two types of scenarios: low-reliability scenarios and high-reliability scenarios. In a high-reliability scenario, the recommended routes were likely to include the highest scoring route. In the low-reliability scenarios, the recommended route never included the highest scoring route. Table 3 shows the rank of the recommended route's score (*1* = highest scoring route).

Table 3. Rank of the recommended route's score (out of 6).

Number of Recommended Routes	Low Reliability	High Reliability
1	2, 2, 3, 3, 3	1 1, 1 ,1, 2
3	2-4, 2-4, 3-5, 3-5, 3-5	1-3, 1-3, 1-3, 1-3, 2-4

2.2.5 Secondary Tasks

The participants also performed two secondary tasks, a monitoring task and a communicating task, that were inspired by the types of tasks performed by traffic managers. There are many aspects to a traffic manager's job and a variety of tasks they need to perform. These secondary tasks were added to simulate some of the demands on the traffic manager's time and attention so we could perform our evaluation of decision support tools in more operationally realistic circumstances.

We varied the frequency of the secondary tasks during the scenario to vary the workload of the participants. In low workload scenarios, each secondary task required 4 responses, while in the high workload scenarios, each task required 10 responses.

2.2.5.1 National Airspace Monitoring Task

This secondary task required participants to monitor and report congestion values from a simulated National Airspace System (NAS) monitor shown in Figure 3. At the start of each scenario, all of the cells were randomly set to green or yellow. Periodically throughout the 4 minutes of rerouting, a cell would change from green or yellow to red (similar to what a traffic manager would see if the Monitor Alert Parameter (MAP) value was exceeded). When a cell turned red, participants were to select the red cell, revealing the predicted aircraft count (18-24) in that sector. Next, the participants had to report these three values: sector, time, and aircraft count by typing the values into three text boxes. In Figure 3, the goal was to select the red cell, enter 14 under “Sector”, enter 2130 under “Time”, and the “Sector Count” number from that cell (not shown), and press the Submit button. The cell remained red for the duration of that scenario. The sector count remained visible to indicate which cell had already been selected.

We were concerned that participants would not pay enough attention to the secondary tasks and only focus on the primary rerouting task. To encourage vigilance on the NAS monitoring task, participants were awarded or penalized points based upon the speed and accuracy of their submissions:

- A correct response within 20 seconds of the cell turning red (+50)
- A correct response, but after 20 seconds (-50)
- An incorrect response (-100)
- No response before scenario end (-100)

Since there could be multiple cells turning red at one time, the task involved a working memory component, in that participants had to remember which of the red cells they were intending to respond to as they completed their submission. This led to an uncertainty regarding which cell was the intended target of an incorrect response since a submission could contain elements belonging to 2 or 3 different red cells. Thus, we knew whether the participants responded to any given cell correctly (all 3 submission fields matched a currently red cell) but we could not definitively distinguish between errors of omission (no submission) versus error of commission (a submission of incorrect information).

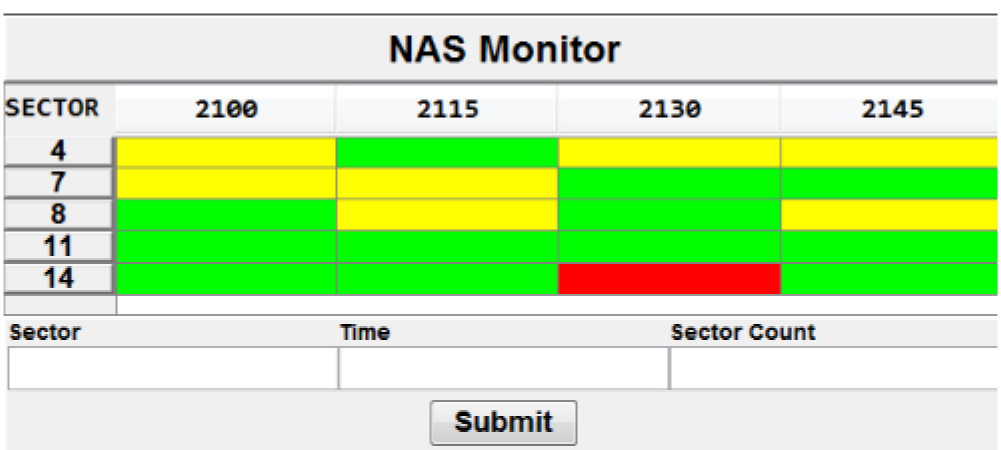


Figure 3. The National Airspace monitoring interface.

2.2.5.2 National Traffic Management Log Communicating Task

This task required participants to monitor the simulated National Traffic Management Log (NTML) window for new messages (see Figure 4). Each message consisted of the time of the message, the message content, and the sender's initials. The message content was based upon content of actual messages that may be seen by traffic managers in the NTML. Each message had to be either "Forwarded" to their supervisor or simply "Acknowledged" (ACK button in see Figure 4). If the message made a reference to either airport involved in the scenario (JFK/ORD or SFO/DEN), it was to be forwarded. If it did not, it was to be acknowledged. To avoid confusion, no messages in the currently active scenario contained reference to the airports used in other scenarios (e.g. in a JFK/ORD scenario, no messages contained SFO or DEN). To incentivize quick and accurate responses, the following points were awarded or deducted:

- A correct response within 10 seconds of message appearing (+50)
- A correct response, but after 10 seconds (-50)
- An incorrect response (-100)
- No response before scenario end (-200)

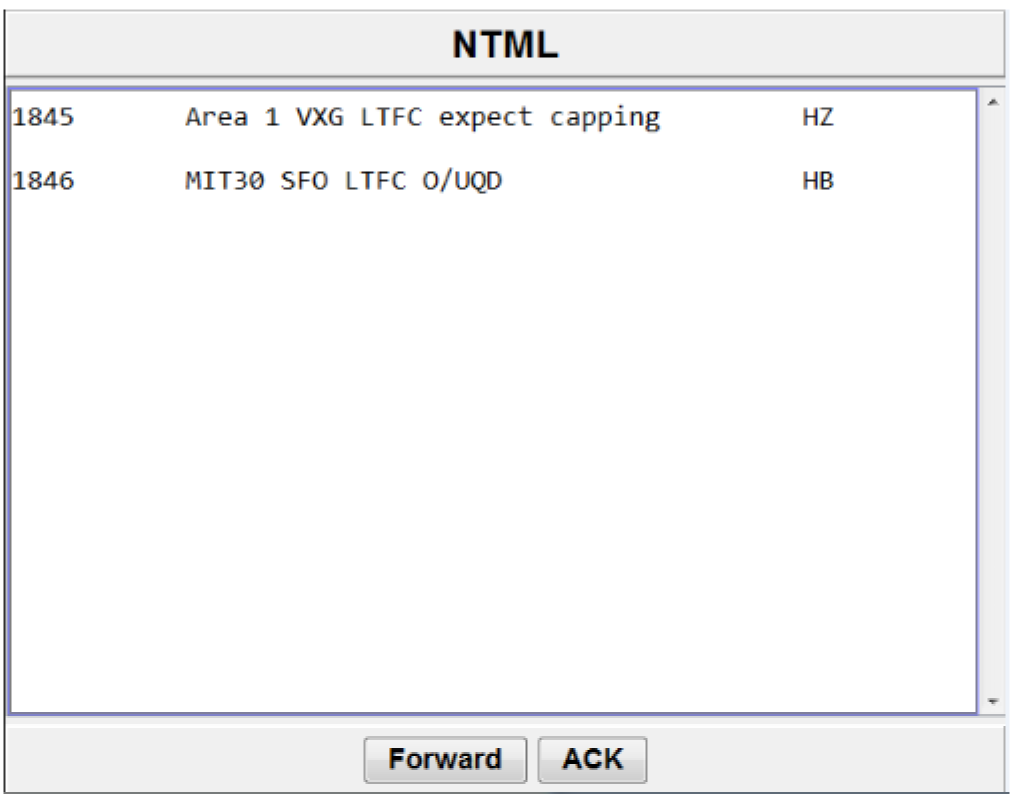


Figure 4. The National Traffic Management Log communicating interface

2.2.6 Feedback

At the conclusion of each scenario, and after completing a short questionnaire, the participants received feedback on their performance in the scenario they just completed. The goal of the feedback was to show the participants how well they did in relation to the automation. The scores for each route for each flight were displayed on the screen. The route that the participant chose was highlighted in yellow. Any routes that would have yielded a higher score than the one chosen were also partially highlighted. All routes that were recommended by the automation were also indicated. The participants could quickly learn if their rerouting choices were optimal, how good the automation's route recommendations were, and if their own choices were better or worse than the automation's suggestions.

The participants were also shown any points accumulated or lost based on their performance on the secondary NAS monitoring and NTML communication tasks. We hoped that displaying the points earned would motivate our participants to be efficient and effective in performing these secondary tasks.

2.2.7 Training

All participants received a training slideshow followed by 5 practice scenarios. The training materials were inspired by the computer-based instruction (CBI) materials reviewed in the TFM Decision Support Tools Training Tools Assessment document (Zingale, Masalonis, Woroch, & Yuditsky, 2017). Participants in both Phase I & II viewed a slideshow containing descriptions of the task. Phase I had an interactive training component, named "CBI-Do", as described in the excerpt below from Woroch, Zingale, and Masalonis (2017):

> "Prior to starting the experimental scenarios, all the participants completed a training session that instructed them how to perform the task and obtain a high score. They also completed practice scenarios to gain experience working on the tasks and to mitigate learning effects in the first few experimental scenarios. The training consisted of a slideshow and interactive demo using the experiment software (the demonstration scenarios were simply modified experimental scenarios with no time component). The slides described the goals of each task and how to perform all elements. After reading several slides, the participants would perform an action or complete an element of the task in the demonstration scenario before reading several more slides in the training slideshow. The first demonstration scenario had no automation suggested routes and guided the participant through the primary rerouting task and both of the secondary tasks. This was followed with a detailed feedback, demonstrating how they performed as well as a verbal description on how to achieve a higher score."

We created a new training slideshow for Phase II. This slideshow contained the same information that we presented to participants in Phase I. However, it did not have the interactive component. Instead of interactions with the experimental software, the Phase II slideshow had screenshots of the software with animation highlighting the corresponding text on the screen. This approach was closer to many of the currently deployed CBI materials. We named this training approach, "CBI-View."

Neither of the groups analyzed in Phase II received the additional SST used in Phase I. The SST contained information about how to predict the reliability of the automation's recommendations. It is not information that is commonly found in current CBI materials, so we omitted it from the current study.

2.3 Procedure & Experimental Design

We included four experimental parameters or independent variables (IVs). We manipulated these variables in our experiment to see the effect they would have on participants' performance of the tasks, known as the dependent variables. There were ten experimental scenarios. We counterbalanced the IVs so that they co-occurred equally. Instead of randomizing condition orders or systematically varying them in a Latin square or similar design, we explicitly specified certain orders for the combinations of independent variables in a way that attempted to avoid overly biasing any participant into a certain initial attitude about the automation or the task.

2.3.1 Workload

We manipulated the participants' workload by the number of events in the secondary tasks during a scenario. There were two levels of workload in this study. In the low workload scenarios, four NAS monitor cells turned red and four NTML messages appeared throughout the scenario. In the high workload scenarios, 10 of each secondary event occurred. The two levels were informed by pilot testing with developmental personnel and other researchers. Of the 10 scenarios, 5 were low workload and 5 were high workload.

2.3.2 Number or Recommendations (NREC)

Each scenario had zero, one, or three recommended routes for each of the five flights. We varied the number of recommendations (NREC) across scenarios. Four scenarios had one recommendation, four had three recommendations and two scenarios had no recommendations. This variable allowed us to address the question as to what "level" of automation (e.g., single best, several options) is associated with higher performance.

2.3.3 Automation Reliability

There were two levels of reliability in this study. The ranks of the recommended routes (out of 6) are shown in Table 4. There were four low-reliability scenarios and four high- reliability scenarios. The two scenarios that had zero automation recommendations could not be reliable or unreliable, by definition. The combination of all variables is shown in Table 4.

Table 4. The ten experimental scenario types performed by all participants.

Scenario:	1	2	3	4	5	6	7	8	9	10
Workload	Low	Low	Low	Low	Low	High	High	High	High	High
NREC	1	1	3	3	0	0	1	1	3	3
Reliability	Low	High	Low	High	N/A	N/A	Low	High	Low	High

3. RESULTS

We analyzed the data to evaluate the effects of the IVs on the various dependent variables. The IVs were: Training Group (CBI-Do vs. CBI-View), Workload (Low vs. High), automation reliability (Low vs. High), and Number of Recommendations (NREC) (1 vs. 3). We conducted three principle types of data analyses and used an alpha level of 0.05 to determine statistical significance. The first analyses were T-tests between the training groups (CBI-Do vs. CBI-View). We tested demographic factors and CPRS scores to determine whether any pre-experiment differences were evident between the participants in the CBI-Do group from the Phase I experiment and the CBI-View group from Phase II. The other two analyses involved a mixed model multiple analysis of variance

(ANOVA) with a between group (CBI-Do vs. CBI-View) factor and repeated-measures factors representing the relevant independent variables. We did this to assess the impacts of all IVs in the same statistical model rather than perform tests of each IV individually and inflating our alpha rate. We used the mixed-model to combine a categorical variable (training group membership) with the repeated-measure factors performed by all participants.

We analyzed the behavioral effects of the IVs using a 2x2x2x2 (Training group x NREC x Workload x Reliability) ANOVA. The dependent variables we analyzed were the scores on the rerouting task, the response times to the secondary tasks, and responses to the survey questions. The rerouting score was the total score obtained by rerouting all five flights in a scenario. We also performed the same analyses on the raw scores reported below as on the proportion of the score obtained for a scenario out of the maximum possible score for that scenario (to normalize the data) The results from those two sets of tests did not differ, so we only report the raw scores.

3.1 Between-Group Demographics & Complacency Rating Scale

We attempted to match the CBI-Do and the CBI-View groups on age to minimize demographic differences between them. We did this so that any differences in performance we observed between the training groups could more likely be attributed to the training manipulation and not to a pre-experimental age difference. There were eight participants (2 female) in the CBI-Do group, age (M = 42.62, SD = 12.39) and eight participants (1 female) in the CBI-View group (M = 38.35, SD = 12.15) The CBI-Do and CBI-View groups did not differ in age, [$t(15) = 0.71$, p = 0.487]. We provide a graph of all 16 participants' ages in Figure 5.

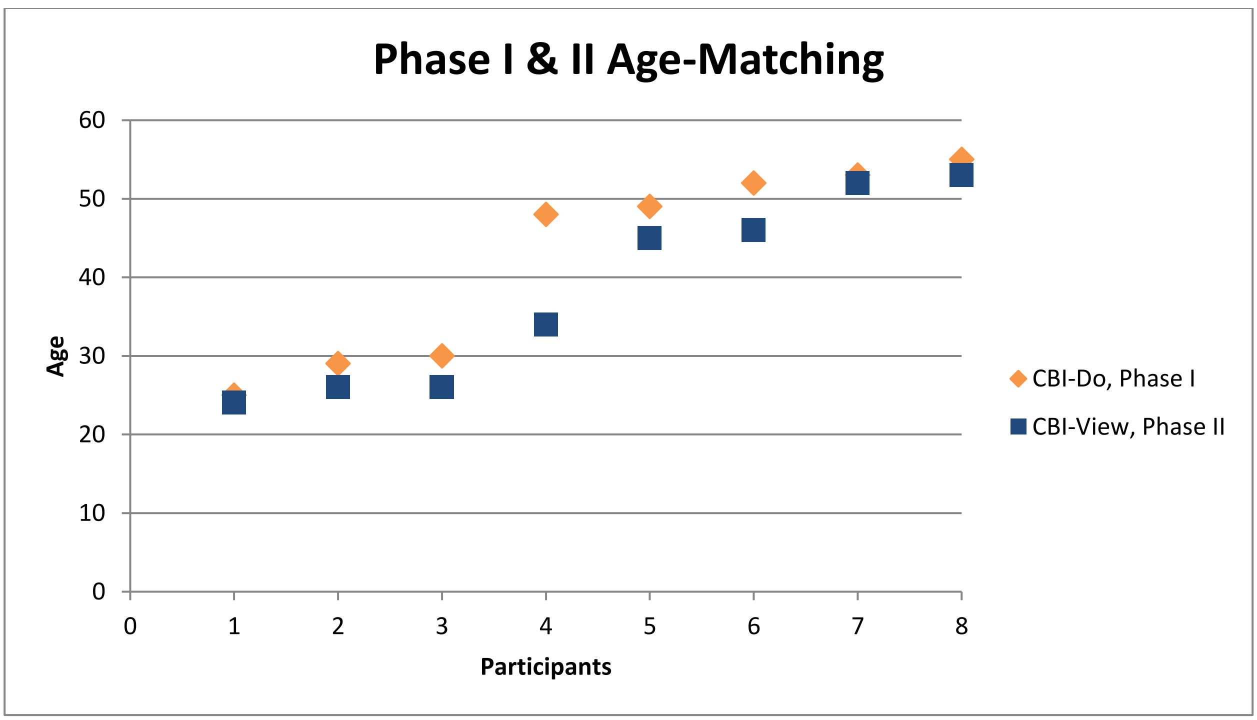


Figure 5. The ages of every participant used in the analysis.

We also found no difference in the Complacency Rating Scale scores between the CBI-Do (M = 62.75, SD = 5.80) and CBI-View (M = 65.13, SD = 6.29) training groups, [$t(14) = 0.78$, p = 0.445]. This gave us further confidence that any observed difference in performance between the CBI-Do and CBI-View groups would be due to the experimental training manipulation and not due to pre-experimental differences between the groups. We assumed our samples were statistically matched and treated them as having equal variance for the purposes of statistics.

3.2 Primary Rerouting Task

The primary dependent variable of interest was the total score from the rerouting task. We obtained one total rerouting score per scenario by adding the scores from the five flights. The range of possible scores in each scenario, with no time penalties, was between 2895 and 4425. In this study, the actual scores per scenario were much higher, ranging from 3975 to 4325.

Two of ten scenarios had no automation recommendations, one for each workload condition. We used a 2x2 (group x workload) ANOVA to analyze the scores for those scenarios. There was no statistically significant difference in score between the CBI-Do and CBI-View groups during the two no-automation scenarios (see Figure 6), [$F(1,14) = 0.25$, $p = 0.623$]. This result also indicated that the two groups did not differ from one another before we introduced the experimental variables.

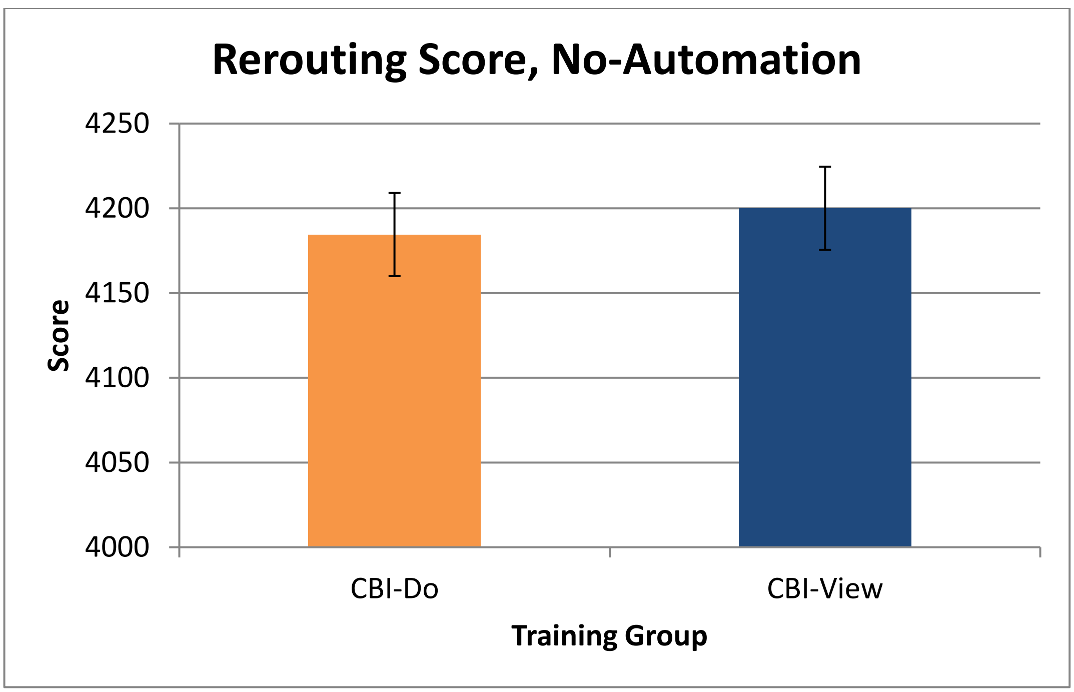


Figure 6. The average rerouting scores in the no-automation condition. The error bars represent the standard error of the mean.

Eight scenarios for each participant featured automated recommendations. The five flights in each scenario were summed and submitted to a 2x2x2x2 ANOVA to test the effects of each independent variable. We analyzed the variables of Workload, Reliability, and NREC in depth for the Phase I report (Woroch, Zingale, and Masalonis, 2017) and we examined them in depth for Phase II. We present the statistically significant results from both Phases I & II in Table 5. The two tests are not independent as they both utilize 8 participants from Phase I. This report will only focus on the results that differ from Phase I and any interaction between the variables and the Training variable.

Table 5. The ANOVA results for Phases I &II

Independent Variable	**Phase I**	**Phase II**
Workload x Reliability	$F(1,14) = 7.24$, $p = 0.018$	$F(1,14) = 3.9$, $p = 0.069$ (N.S.)
NREC x Reliability	$F(1,14) = 11.07$, $p = 0.005$	$F(1,14) = 8.20$, $p = 0.013$
Workload	$F(1,14) = 52.77$, $p < 0.001$	$F(1,14) = 67.88$, $p < 0.001$
Reliability	$F(1,14) = 43.63$, $p < 0.001$	$F(1,14) = 11.70$, $p < 0.001$
NREC	$F(1,14) = 0.07$, $p = 0.796$	$F(1,14) = 4.4$, $p = 0.055$ (marginal)

The pattern of results between the two phases was similar, with two exceptions: the interaction between workload and reliability observed in Phase I was no longer statistically significant in Phase II. Additionally there was no main effect of the NREC in Phase I, but it was marginally significant in Phase II. We did find a significant interaction between the Training group (CBI-Do and CBI-View) and Workload [$F(1,14)=10.7$, $p = 0.006$] with the CBI-View group outperforming the CBI-Do group in scenarios of high Workload as shown in Figure 7.

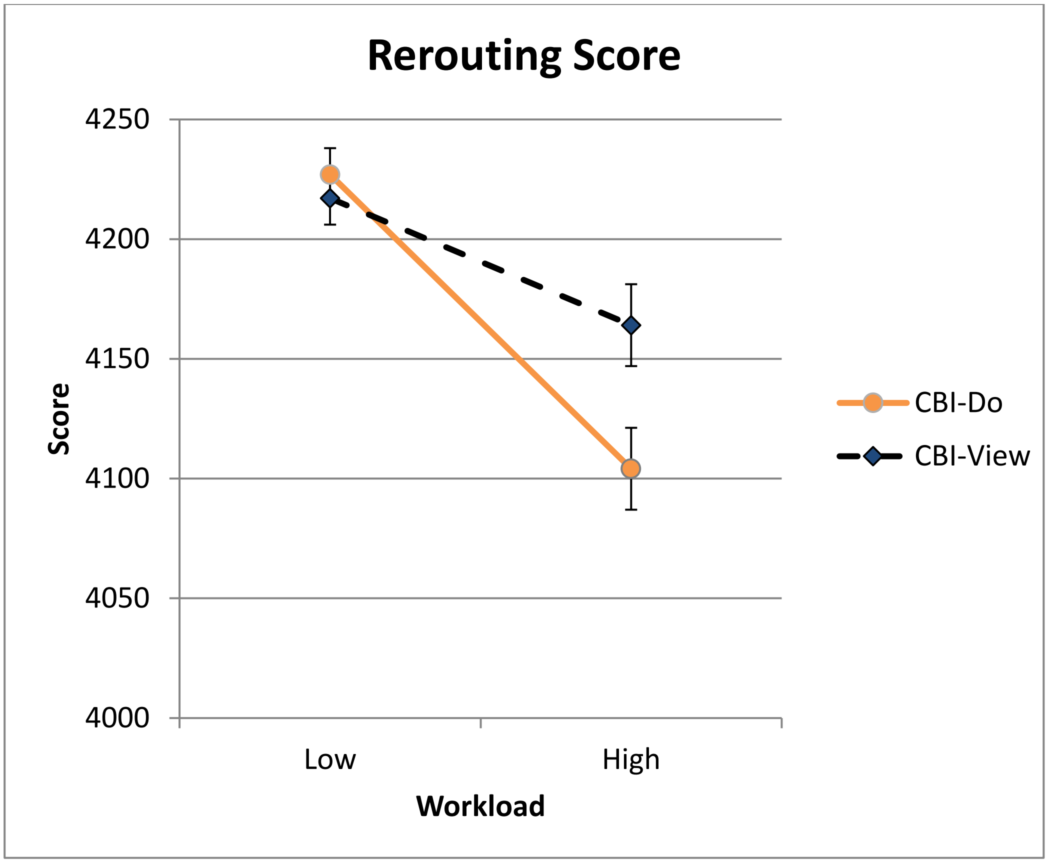


Figure 7. Interaction of Workload and Training Group on rerouting performance. The error bars represent the standard error of the mean.

As we found in Phase I, a lower task Workload yielded better performance overall, as did higher automation Reliability. Figure 8 is a graph of the mean score for each of the independent variables.

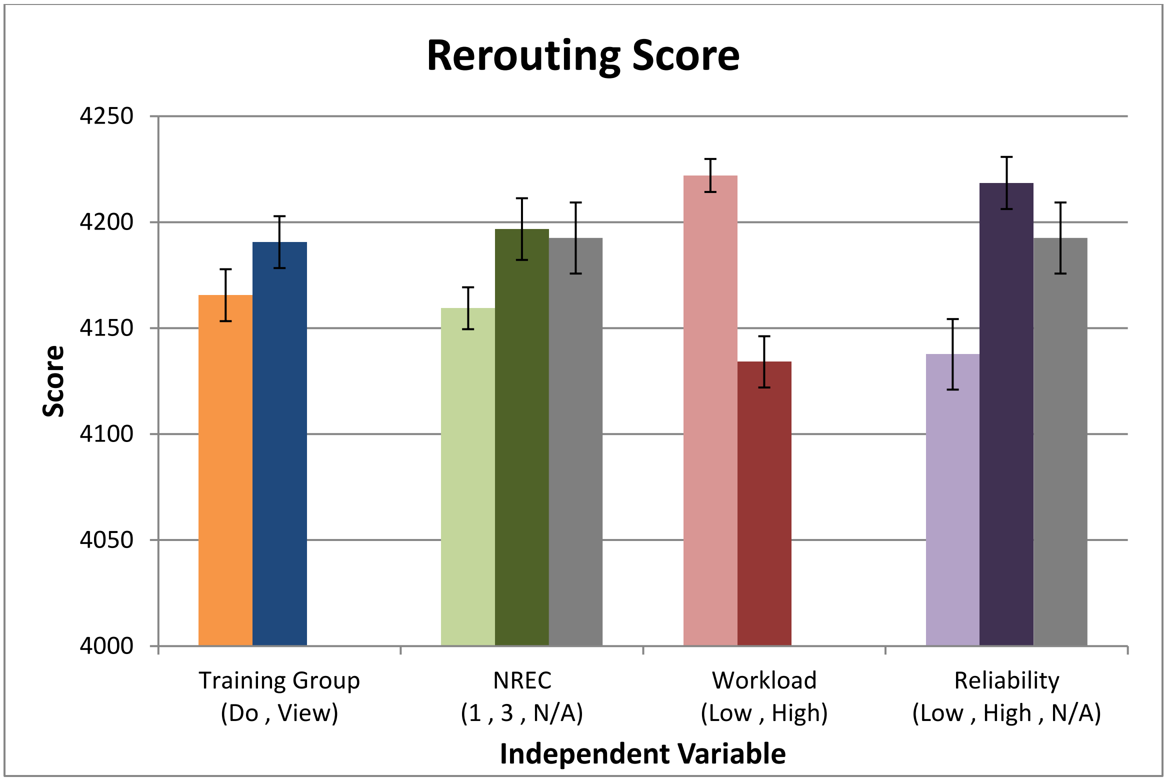


Figure 8. The average rerouting score in each condition. The error bars represent the standard error of the mean.

We did not find a main effect of CBI Training method on rerouting score [$F(1,14) = 2.1$, $p = 0.170$], the primary comparison of interest. We decided to examine how learning the task unfolded over time to see if any interesting differences between the Training groups emerged. We included the practice scenarios in this assessment. Figure 9 shows the average rerouting score for each scenario in chronological order. The practice scenarios (P1 – P5) were the same length as the experimental scenarios (1-10). All participants performed the same five practice scenarios. The experimental scenarios were counterbalanced across participants to be equally likely to be high or low workload, 1 or 3 recommendations, or high or low reliability (exception: Scenarios 1 & 2 or 9 & 10 were no-automation scenarios).

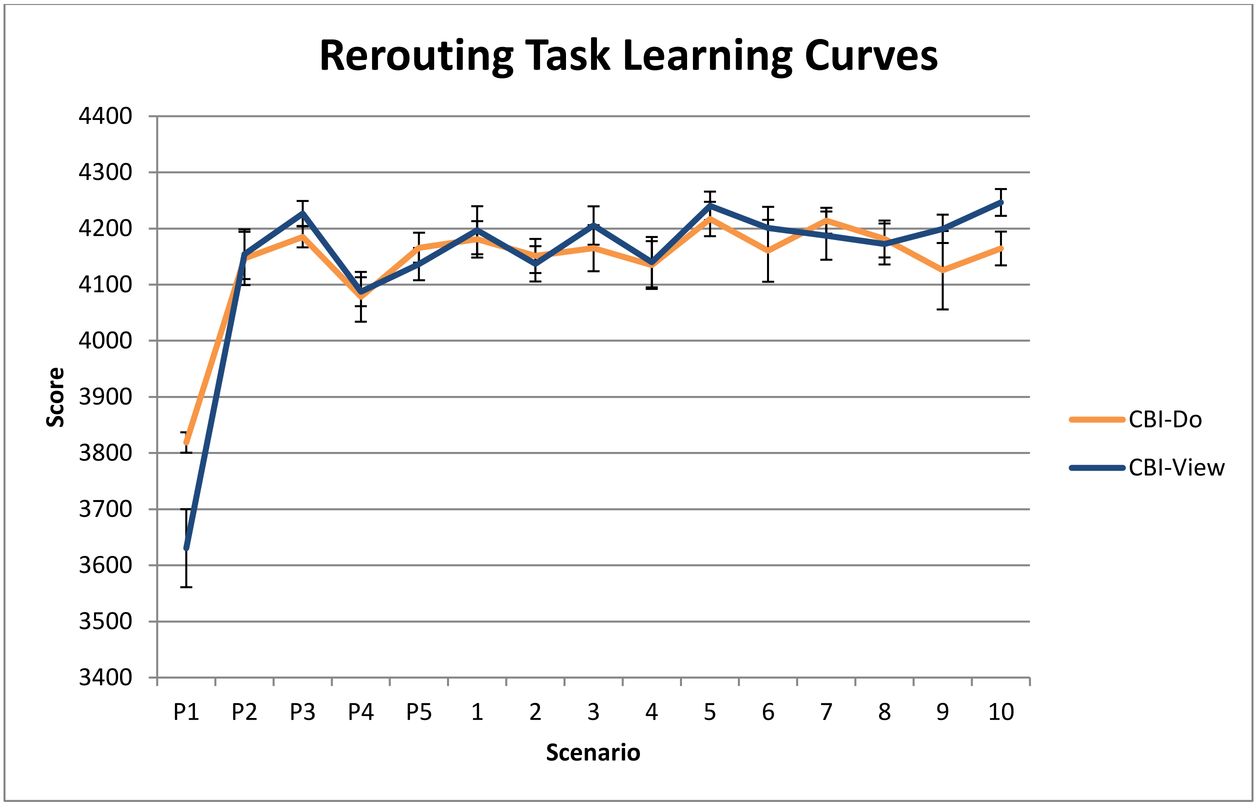


Figure 9. The average rerouting score for each scenario, practice (P1-P5) & experimental (1-10). The error bars represent the standard error of the mean.

We conducted post-hoc tests on each data point. In the first practice scenario, the CBI-Do group out-performed the CBI-View group [$t(15) = 2.59$, $p = 0.020$]. However, by the second practice scenario, the performance of the two groups was equivalent and remained the same until the final scenario [all $t(15) < 1.43$, all $p > 0.172$]. We feel this result is meaningful, and we will discuss the potential implications and useful next steps in the Summary and Recommendations section. In the final experiment scenario, the CBI-View group performed slightly better, just above a $p < 0.05$ threshold [$t(15) = 2.12$, $p = 0.05083557$]. It should be noted that these p values are uncorrected for multiple comparisons and, therefore have an inflated type-I error rate. These results should be viewed as preliminary. Follow-up research is needed for verification.

3.3 Secondary Task Results

In addition to the primary rerouting task, participants completed two simultaneous secondary tasks. We attempted to analyze the data from these tasks using the same ANOVA models used for the primary rerouting task, and only report effects that were significant at $p < 0.05$. As we found for Phase I, the accuracy for both secondary tasks in Phase II was extremely high (>90%) and a lack of variance in some conditions meant that accuracy could not be analyzed with the ANOVA model. However, once again, we were able to use response time to these tasks as a dependent variable.

3.3.1 NAS Monitoring & Reporting

The NAS Monitoring and Reporting task required the participants to acknowledge when a sector turned red. The participants selected the "red" cell and entered the number of aircraft expected in a text box. The mean accuracy for this task for the CBI-View group from Phase II was 99.11% (0.013%). The ANOVA model could not be fit to the accuracy data since some conditions had no variance (all participants scored 100%). Direct comparisons between the means of each condition using a T-test revealed no differences in accuracy between conditions. However the CBI-View group did outperform the CBI-Do group from Phase I ($M = 96.33\%$, $SD = 3.42\%$) [$t(15) = 2.26$, $p = 0.039$].

The other dependent variable of interest for this task was the response time (RT) between a cell in the NAS array turning red and the participant selecting it to reveal the flight count (see Figure 3 from section 2). We found no difference in RT between training groups [$F(1,14)=0.76$, $p = 0.400$]. Similar to Phase I, NAS Monitoring RT was faster when workload was lower [$F(1,14) = 10.06$, $p = 0.007$]. There were no other main effects and no two-way interactions. However, there was a significant three-way interaction of NREC x Reliability x Training group [$F(1,14) = 4.72$, $p = 0.048$]. In the absence of other main effects and two-way interactions, interpretation of the functional significance of the three-way interaction is difficult to interpret. The functional significance of this effect is unclear.

3.3.2 NTML

The other secondary task pertained to NTML messages. The participants had to either acknowledge or forward each message as described in section 2.2.5.2. The dependent variable of interest for the NTML communication task was the RT between a message appearing in the NTML window and the participant forwarding or acknowledging the message.

There was no difference in NTML RT between the training groups [$F(1,14) = 0.08$, $p = 0.783$]. Just as in Phase I, the only significant effect on NTML RT was workload. Response time was faster when workload was lower [$F(1,14) = 11.98$, $p = 0.004$].

3.4 Task Questionnaire Results

We presented the same three questionnaires to the participants during this experiment as we did in Phase I. All questionnaires used 10-point scales with *1* indicating the lowest rating and *10* indicating the highest rating. We analyzed the surveys in depth for the Phase I report (Woroch, Zingale, and Masalonis, 2017). The results for this study followed the same pattern of results as Phase I. Of interest to this study were any differences in survey responses between the training groups or significant interactions with training. However, we did not find any significant differences.

The survey response after each reroute once again showed that participants rated their own performance higher under low Workload and relied on the automations recommendations more during high Reliability scenarios. The survey response after each scenario showed a similar pattern. Participants rated their workload as higher under high Workload conditions and the RRT's performance higher in high Reliability scenarios. The end of experiment survey showed no difference between groups to any questions.

Phase II also included a questionnaire at the end of the experiment regarding the experimental software usability and the effectiveness of the training. We present the descriptive results in Table 6 The median ratings for these items was high. All of the median ratings were seven or higher, indicating indicated that the participants found the software easy to use and training effective.

Table 6. The scores from the usability and training surveys.

Question	Median	Range
1. The software was easy to use.	9	8-10
2. The displays were uncluttered.	9.5	7-10
3. I could find the information I needed quickly.	8.5	7-10
4. I was able to prioritize information and tasks easily.	7	6-9
5. It was easy to notice changes or updates to the display.	7.5	5-9
6. The information was presented in an easy-to-understand format.	9	7-10
7. It took only a few simple steps to get the information I needed when it wasn't available directly on the display.	9	5-10
8. The coding of the information (e.g., colors) helped me locate what I needed quickly.	9	4-10
9. The text was easy to read.	9	7-10
10. The graphics were easy to interpret.	8.5	7-10
1. The training slideshow was easy to understand.	9	5-10
2. The training slideshow taught me everything I needed to perform the tasks.	9	5-10
3. The practice scenarios were easy to complete.	9	5-10
4. The amount of practice was enough to prepare me for the experiment scenarios.	9	3-10

4. SUMMARY

We conducted the current experiment to evaluate the effects of different types of training on DST effectiveness. We matched the demographics of the eight Phase II participants with the existing eight participants from Phase I to minimize pre-experimental differences between the groups. There were no differences in the ages of the two groups. The two groups also scored similarly on the Complacency Rating Scale, a measure of susceptibility to overreliance on automation.

Both groups of participants performed the same experimental scenarios that consisted of a primary aircraft rerouting task and two secondary tasks. We found that the training method had no effect on the overall aircraft rerouting performance in the experimental scenarios. Although we hypothesized that the CBI-View training method would lead to impoverished task performance compared to the CBI-Do method, we did not find evidence to support this hypothesis. However, we did notice some interesting differences between the CBI-Do and CBI-View Groups. The CBI-View group performed the rerouting task better in scenarios with a high workload and had higher overall accuracy on one of the secondary tasks.

Although, we attempted to match the CBI-View group to the CBI-Do group closely, it is possible that the new group of participants happened to be better at this task than the previous group of participants. Their improved performance might have compensated for any deficit in training. It might also be that our experimental task was too easy to learn and that the CBI-View method was fully effective in teaching this task.

Interestingly, we did observe that the CBI-View group performed the rerouting task worse on the very first practice scenario, and we believe that this has some important implications. First, we will discuss some of the differences that may also have influenced performance, between the CBI training method we developed and the existing TFM CBI training materials we reviewed.

4.1 Differences between the CBI-View method and operational CBIs

There were some differences between our CBI-View materials and the CBI materials we reviewed in the Training Tools Assessment document (Zingale, Masalonis, Woroch, & Yuditsky, 2017). These differences may have increased the effectiveness of our training in the current experiment. We have identified two aspects we feel are important. These differences represent concepts that could improve current training methods

4.1.1 Self-Pacing

We implemented our training in a self-paced slideshow format. The participants proceeded at the rate they chose. They could quickly go through slides that presented simple ideas and take longer to read and understand slides with more complex concepts. This is in contrast to many of the current CBI packages. Those slideshows often include narration, and the user cannot control the rate of information presented. The ATC subject-matter experts (SMEs) who evaluated many of the materials with us noted that some slides moved much too fast for them. At times, they were overwhelmed with new information and did not have time to understand new concepts before more information was presented.

If the participants in our study wanted to repeat a slide or a single piece of information, they could easily navigate back and forth through the slides. In a typical CBI, if users want to review a single point, they often need to navigate back through a main menu and repeat an entire section of the material until the piece of information they were looking for is presented.

We feel all CBIs would be more effective if they allowed the user to self-pace and provided easier navigation. A preset pacing assumes all users will understand each topic in the same amount of time. What may be easier for some to understand may be more difficult for others. Allowing the user to set the pace means time is not wasted on simple points and that more time can be allocated to complex concepts. Leaving the rate up to the user ensures that individual differences in learning and comprehension are accommodated.

4.1.2 Feedback

In these studies, each scenario (including practice), was followed by feedback showing the participants how they performed. This feedback may have been instrumental in helping participants learn how to perform the task effectively. Participants in the CBI-View group performed more poorly than the CBI-Do group on the first practice scenario. However, the CBI-View group performed just as well as the CBI-Do group in the second practice scenario. The feedback the participants received after the first practice scenario may have enabled the CBI-View group to use the software more effectively.

4.1.3 Implications

This study required participants to continue learning and practicing the rerouting task. However, in an operational setting, the use of a new tool may be optional. The first time users encounter a new tool is in training. If the initial use of a new tool yields higher performance, it may lead to a more positive first impression and higher perceived usefulness. This positive impression may then lead to continued use following training. If the initial use of a new tool leads to poor performance, a poor first impression, and decreased perceived usefulness, the tool may not be used, or may be under-utilized following training. Research from the information technology field has shown that perceived usefulness is positively correlated with usage of new hardware and software (Davis, 1989; Adams, Nelson, & Todd, 1992). Although our study does not provide direct evidence that initial performance is correlated with the initial impression and perceive usefulness, we would hypothesize that it is. Future research is needed to directly test this hypothesis.

Based upon the existing data and our knowledge of learning and memory, we recommend that training with CBIs include practice opportunities for users. These practice opportunities should be followed by feedback about performance and demonstrate how to improve performance to overcome any initial negative impressions and increase perceived usefulness.

5. CONCLUSIONS & RECOMMENDATIONS

This study tested whether the interactive training (CBI-Do) resulted in better task performance than a more passive form of training (CBI-View). We did not find evidence that the interactive training method improved task performance on the experimental scenarios. However, we did find that interactive training led to increased performance on the initial practice scenario of the task. This benefit may have implications for real-world training. If the first impression of a new tool is positive and its perceived usefulness is high, it may lead to increased adoption and future use. However, if the first impression of a new tool is negative, or if the tool is difficult to learn, users may abandon it or use it less often. Future research on this topic would be informative.

Our non-interactive training also implemented two features we recommend as useful additions to current operational CBIs. We recommend that training materials:

- Allow the user to set the pace for learning instead of requiring that the user proceed at the pace imposed by the automation.
- Provide the user specific feedback about performance and demonstrate the potential benefits of new tools and procedures.

References

Adams, D. A., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly*, 227-247. https://doi.org/10.2307/249577

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319-340. https://doi.org/10.2307/249008

Davison Reynolds, H. J., & DeLaura, R. A. (2011). *Concept of operations for the Integrated Departure Route Planning (IDRP) tool* (Project Report No. ATC-379). Lexington, MA: Lincoln Laboratory, Massachusetts Institute of Technology.

DeLaura, R. A., Underhill, N. K., Hall, L. M., & Rodriguez, Y. G. (2012). *Evaluation of the Integrated Departure Route Planning (IDRP) tool 2011 prototype* (Project Report No. ATC-388). Lexington, MA: Lincoln Laboratory, Massachusetts Institute of Technology.

Means, B., Mumaw, R. J., Roth, C., Schlager, M. S., McWilliams, E., Gagne, E., & Heon, S. (1988). *ATC training analysis study: Design of the next-generation ATC training system*. FAA/OPM, 342-036.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the Complacency-Potential Rating Scale. *International Journal of Aviation Psychology, 3(2)*, 111–122. https://doi.org/10.1207/s15327108ijap0302_2

Woroch, B., Zingale, C.M., & Masalonis, A.J. (2017). *Traffic Flow Management Tools: Guidance for Use, Integration, and Training: Part-Task Experiment 1*. DOT/FAA/TC-17/37. Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Zingale, C.M., Masalonis, A.J., Woroch, B., & Yuditsky, T. (2017). *Traffic Flow Management Tools: Guidance for Use, Integration, and Training: Phase II – Human Factors Assessment of Computer and Web-based Training for Decision Support Tools*. DOT/FAA/TC-17/36. Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Acronyms

ARTCC	Air Route Traffic Control Center
ATC	Air Traffic Control
ATCSCC	Air Traffic Control System Command Center
ATCT	Air Traffic Control Tower
DST	Decision Support
IV	Independent Variable
NAS	National Airspace System
NREC	Number of Recommendations
NTML	National Traffic Management Log
RRT	Route Recommendation Tool
SST	Situation Specific Training
TFM	Traffic Flow Management
TLX	Task Load Index
TRACON	Terminal RADAR Approach Control

Appendix A: Informed Consent Statement

Informed Consent Statement

I, ___________________________________, understand that this study, entitled NextGen Traffic Flow Management (TFM) Tools: Guidance for Use, Integration, and Training, is sponsored by the Federal Aviation Administration (FAA).

Nature and Purpose:

I have been recruited to volunteer as a participant in this project. The purpose of this study is to develop a better understanding of human behavior when using the types of Decision Support Tools planned for the Traffic Flow Management domain. The results of this study will generate considerations and suggestions for adding Decision Support Tools in the TFM environment.

Study Procedures:

Approximately sixteen (16) volunteers, primarily from the William J. Hughes Technical Center, will participate in this experiment, designed to simulate some of the demands of TFM. The experiment plus all training and questionnaires will take approximately 2 hours. After the conclusion of the experiment, participants and researchers will conduct a final debriefing session to share questions, comments, and feedback.

Anonymity and Confidentiality:

My participation in this simulation is strictly confidential. Any information I provide will remain anonymous; no individual names or identities will be associated with the data or released in any reports.

Benefits:

I understand that the only benefit to me is that I will be able to provide valuable feedback and insight into the effectiveness of potential ATC tools and procedures. My contribution will help the FAA to determine the benefits and feasibility of these modifications.

Participant Responsibilities:

I will perform the tasks presented during this study to the best of my ability and will answer all questions asked during the study truthfully. I will not discuss the content of the study with anyone until the study is completed.

Participant Assurances:

I understand that my participation in this study is completely voluntary and I can withdraw at any time without penalty. I also understand that the researchers in this study may terminate my participation if they feel this to be in my best interest. I understand that if new findings develop during the course of this study that may relate to my decision to continue participation, I will be informed. I have not given up any of my legal rights or released any individual or institution from liability for negligence.

The research team has adequately answered all the questions I have asked about this study, my participation, and the procedures involved. I understand that Carolina Zingale or another member of the research team will be available to answer any questions concerning procedures throughout this study. If I have questions about this study or need to report any adverse effects from research procedures, I will contact Carolina Zingale at 609-485-8629.

Discomfort and Risks:

I understand that I will not be exposed to any known risks or intrusive measurement techniques. I agree to immediately report any injury or suspected adverse effect to Carolina Zingale.

Signature Lines:

I have read this informed consent form. I understand its contents, and I freely consent to participate in this study under the conditions described. I understand that, if I want to, I may have a copy of this form.

Participant:____ Date:____________

Investigator:____ Date:____________

Witness:____ Date:____________

Appendix B: Demographic Questionnaire

Demographics Questionnaire

Participant #___________ Age _______________ Gender________________

Do you have normal color vision? (Y / N)

If no, please explain:____

Have you ever worked as an Air Traffic Controller? (Y / N)

____

Have you ever worked as a Traffic Manager? (Y / N)

____

Are you or have you ever been a licensed pilot? (Y / N)

If yes, please indicate license and experience____

What other experience (e.g., R&D) do you have with Air Traffic Control, aircraft piloting, or airline operations?

____

____

____

What experience (e.g., R&D) do you have with Traffic Flow Management?

____

____

____

Have you discussed this experiment with any of your colleagues who have already participated in it?

YES / NO

What do you know about the tasks you will be performing?

____

Appendix C: Modified Complacency Rating Scale

Modified Complacency Rating Scale

INSTRUCTIONS

Read each statement carefully and check one response out of five alternatives in the appropriate box which you feel most accurately describes your views or experiences. The responses vary on a scale of agreement/disagreement, from “strongly agree” to “strongly disagree”. For example:

Statement: Doing research in a library has been made easier by the introduction of computerized card cataloging systems.

☐	☐	☐	☐	☐
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

Give your answer for each statement and be sure to place your response in the correct place. Remember, this is an opinion survey and not a test of intelligence or ability. There are no right or wrong answers, only answers that fit your views accurately. Do not skip any question. Time is limited. Do you have any questions?

1. Manually sorting through card catalogues is more reliable than computer-aided searches for finding items in a library.

☐	☐	☐	☐	☐
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

2. If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because computerized surgery is more reliable and safer than manual surgery.

☐	☐	☐	☐	☐
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

3. People save time by using automatic teller machines (ATMs) rather than a bank teller for banking transactions.

☐	☐	☐	☐	☐
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

4. I do not trust automated devices such as ATMs and computerized airline reservation systems.

☐	☐	☐	☐	☐
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

5. People who work frequently with automated devices have lower job satisfaction because they feel less involved in their job than those who work manually.

☐	☐	☐	☐	☐
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

6. I feel safer depositing my money at an ATM than with a human teller.

☐	☐	☐	☐	☐
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

7. I have to tape an important TV program for a class assignment. To ensure that the correct program is recorded, I would use the automatic programming facility on my DVR rather than manual recording.

☐	☐	☐	☐	☐
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

8. People whose jobs require them to work with automated systems are lonelier than people who do not have to work with such devices.

☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

9. Automated systems used in modern aircraft, such as the automatic landing system, have made air journeys safer.

☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

10. ATMs provide a safeguard against the inappropriate use of an individual's bank account by dishonest people.

☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

11. Automated devices used in aviation and banking have made work easier for both employees and customers.

☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

12. I often use automated devices.

☐ Strongly Agree ☐ Agree ☐ Undecided ☐ Disagree ☐ Strongly Disagree

13. People who work with automated devices have greater job satisfaction because they feel more involved than those who work manually.

☐ ☐ ☐ ☐ ☐

Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree

14. Automated devices in medicine save time and money in the diagnosis and treatment of disease.

☐ ☐ ☐ ☐ ☐

Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree

15. Even though the automatic cruise control in my car is set at a speed below the speed limit, I worry when I pass a police radar speed-trap in case the automatic control is not working properly.

☐ ☐ ☐ ☐ ☐

Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree

16. Bank transactions have become safer with the introduction of computer technology for the transfer of funds.

☐ ☐ ☐ ☐ ☐

Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree

17. I would rather purchase an item using a computer than have to deal with a sales representative on the phone because my order is more likely to be correct using the computer.

☐ ☐ ☐ ☐ ☐

Strongly Agree | Agree | Undecided | Disagree | Strongly Disagree

18. Work has become more difficult with the increase of automation in aviation and banking.

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
☐	☐	☐	☐	☐

19. I do not like to use ATMs because I feel that they are sometimes unreliable.

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
☐	☐	☐	☐	☐

20. I think that technology used in medicine, such as CAT-scans and ultrasound, help to provide very reliable medical diagnosis.

Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
☐	☐	☐	☐	☐

Appendix D: Experiment Questionnaires

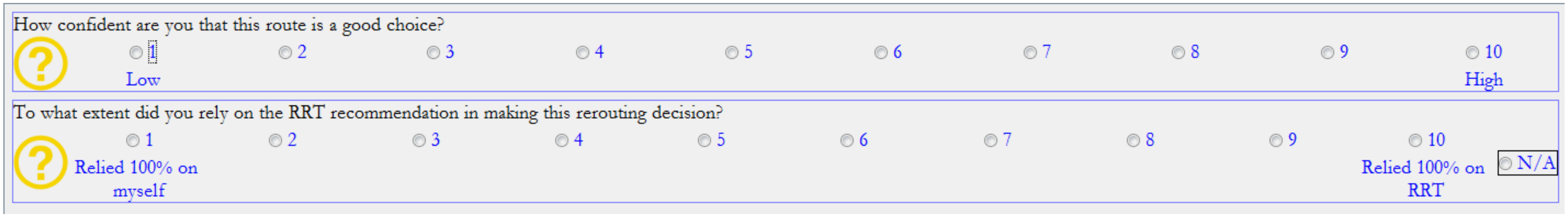


Figure D1. The survey questions presented after each aircraft reroute. N/A was used when there was no automation.

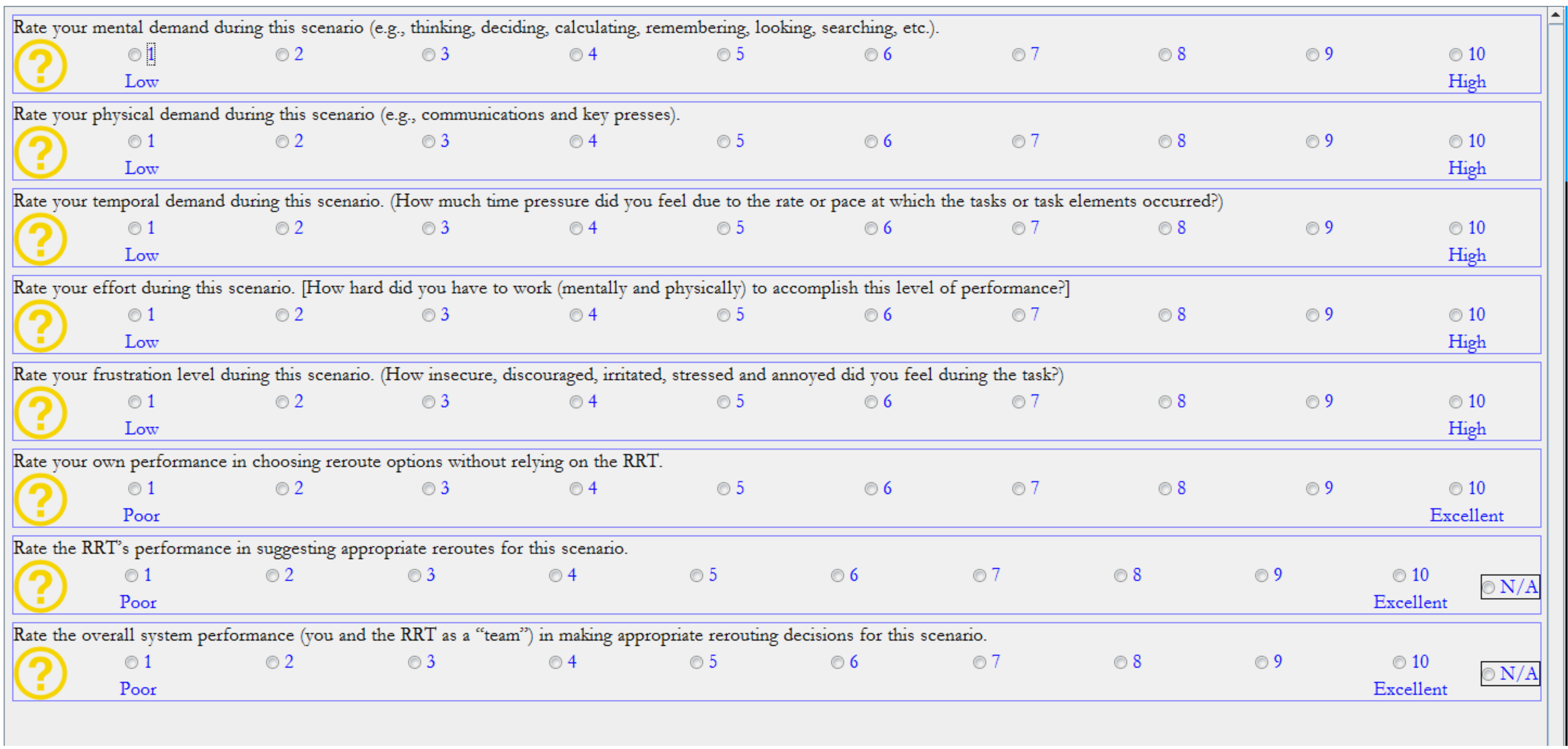


Figure D2. The survey questions presented after each scenario was completed. N/A was used when there was no automation.

Figure D3. The survey questions presented at the conclusion of all 10 test scenarios. N/A was used when there was no automation.

Interface Usability Questions

Rate how much you agree with each of the following statements:

1. The software was easy to use.	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
2. The displays were uncluttered.	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
3. I could find the information I needed quickly.	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
4. I was able to prioritize information and tasks easily.	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
5. It was easy to notice changes or updates to the display.	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
6. The information was presented in an easy-to-understand format.	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
7. It took only a few simple steps to get the information I needed when it wasn't available directly on the display.	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
8. The coding of the information (e.g., colors) helped me locate what I needed quickly.	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
9. The text was easy to read.	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
10. The graphics were easy to interpret.	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**

Additional Comments:

Training & Practice Questions

Rate how much you agree with each of the following statements:

1. **The training slideshow was easy to understand.**	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
2. **The training slideshow taught me everything I needed to perform the tasks.**	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
3. **<u>What would make the training slideshow more effective?</u>**			
4. **The practice scenarios were easy to complete.**	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
5. **The amount of practice was enough to prepare me for the experiment scenarios.**	**Not At All**	①②③④⑤⑥⑦⑧⑨⑩	**A Great Deal**
6. **<u>How could the practice be more effective?</u>**			