UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

REDUCTION OF FLIGHT PROGRESS STRIP ACTIVITY FOR EN ROUTE AIR TRAFFIC CONTROL

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

Todd R. Truitt
Norman, Oklahoma
1999

UMI Number: 9935533

REDUCTION OF FLIGHT PROGRESS STRIP ACTIVITY FOR EN ROUTE AIR TRAFFIC CONTROL

A Dissertation APPROVED FOR THE DEPARTMENT OF PSYCHOLOGY

BY

F T Durso

Robert Terry

Joseph Lee Rodgers

Scott Gronlund

Kirby Gilliland

Robert E. Schlegel

Acknowledgments

This research was supported by grant DOT-FAA 96-G-040 awarded to Dr. Francis T. Durso at the University of Oklahoma.

Table of Contents

Abstract

The new Display System Replacement (DSR) being implemented in air route traffic control centers (ARTCC) will provide controllers less room to post Flight Progress Strips (FPS). The current experiment tested a new FPS marking and posting procedure designed to reduce the controller's need for, or reliance on, the FPS.

The experiment was conducted at Cleveland (ZOB) and Jacksonville (ZJX) ARTCCs utilizing individual controllers and controller teams operating in either high or low altitude sectors. Each participant ran two, 30-minute scenarios. In the Normal condition, participants worked as they normally would. During the Optional FPS condition, participants removed FPSs that were not needed after radar contact and communications were established. Also, FPS marking was not required for any information that was recorded elsewhere, such as via computer entry or landline communication. Scenarios were counterbalanced but sample sizes did not allow counterbalancing of conditions. Participants responded to the Workload Assessment Keypad (WAK) every five minutes while a subject matter expert made performance ratings for each participant or team of participants. Experimenters recorded activity relevant to the plan view display, computer readout device, and FPSs that may be used as a means to compensate for reduced FPS activity. Participants provided a position relief briefing at the end of each scenario and completed a modified version of the NASA Task Load Index (TLX).

For individuals and teams at ZOB and ZJX, participants removed proportionally more FPSs and marked them less often in the Optional FPS condition. There was no effect on performance and participants did not seem to compensate by using flight plan readouts, route displays, or any other means during the Optional FPS condition. On-line measures (WAK) and post-scenario measures of subjective workload (TLX) were comparable for the two conditions. Overall, the Optional FPS procedure appeared to be a viable means by which controllers' reliance and use of the FPS may be reduced.

Reduction of Flight Progress Strip Activity for En Route Air Traffic Control

Currently, en route air traffic controllers use paper flight progress strips (FPS) to provide safe and efficient radar service. The controller uses the FPS to obtain information about a flight and to record changes in flight parameters such as route, speed, or altitude. Controllers are required by Federal Aviation Administration (FAA) procedures to post FPS for all aircraft within their particular sector of airspace. However, much of the information on the FPS can be obtained elsewhere, such as from the computer readout device (CRD). Additionally, much of the information that the controller is required to write on the FPS is also recorded elsewhere, such as on the host computer system or ground-to-air audio tapes. Although redundancy can be very beneficial, particularly when a system fails, redundancy in highly reliable systems may create additional task requirements in terms of workload or cognitive processing for the operator.

The purpose of the current experiment was to examine the performance and workload effects of removing some of the redundant behaviors associated with the FPS that are required for en route air traffic control (ATC), namely, FPS posting and marking. The outcome is of interest due to impending replacement of workstations used by en route controllers. The old, vacuum tube dependent workstation, or M-1 console, is being replaced by the new, more reliable, Display System Replacement (DSR). The DSR was designed to eliminate many of the problems associated with the poorer reliability of the M-1. The DSR has much more computer power to allow for future system upgrades, including the use of color and additional functions for the controller.

One important difference is that the DSR workstation provides less room for FPS management. Therefore, it was reasonable to ask whether the benefits provided by the DSR display and its ability to accommodate new electronic displays and tools would be diminished by a restricted ability to use the FPS. It is important to note that during the initial transition, the DSR will simply replace the old M-1 console without adding any new features. If restricting controller interaction with the FPS results in a deficit (e.g., poorer performance, higher workload) while using the M-1 console, then it is likely that a similar deficit will remain during a transition to the DSR.

A number of researchers have emphasized the importance of active FPS usage (e.g., Hopkin, 1988, 1995; Stein, 1991; Stein & Bailey, 1994; Zingale, Gromelski, & Stein, 1992). Hopkin specifically argued that active control procedures are necessary for controllers to maintain a sufficient level of knowledge and situation awareness (SA) during the ATC task. Hopkin emphasized the importance of physical interaction such as resequencing or writing on the FPS. Without such physical interaction, he argued, controller memory, SA, and hence, overall performance, would suffer. The views of Hopkin and others rest on the ideas that memory encoding is important and that it cannot sufficiently occur without such meaningful physical activity. The importance of memory and its relationship to ATC performance in general, and particularly with regard to FPS usage, is formally acknowledged in a Federal Aviation Administration publication entitled, *The Controller Memory Guide* (Stein & Bailey, 1994). This memory guide has been distributed to controllers in all ATC environments to promote the importance of memory practices and to show how FPS usage can support a possibly overtaxed or undertaxed cognitive system. Field controllers from en route centers, terminal radar control, and airport towers agreed that memory is an important aspect of ATC and rated the memory guide high in terms of relevance, realism, and overall quality (Stein, 1991). Furthermore, interviews of 170 controllers throughout the United States indicated that the three memory aids used most by controllers involve the FPS (Gromelski, Davidson, & Stein, 1992). These often-used memory aids are FPS management (arrangement of FPS), offset or tilted FPS (indication that further action is needed), and FPS marking (updating and confirmation of commands issued).

The importance of physical interaction with the FPS for controller memory and performance has received empirical support in both basic and applied research settings. For example, basic research by Slamecka and Graf (1978) used undergraduates to demonstrate the "generation effect." They demonstrated that participants had better memory for words that were generated by the participants themselves as compared to words that were simply read. The generation effect held under numerous conditions including both recall and recognition memory regardless of whether memory was cued or not.

In a more relevant setting, Zingale, Gromelski, and Stein (1992) provided support for the importance of interaction with the FPS. These researchers trained aviation students to use TRACON II, a simplified, terminal radar ATC simulator. Participants were provided with an FPS for each aircraft in the simulation. Each participant controlled traffic in both Writing and No-Writing conditions in which they either could or could not record control actions on the FPS. Results showed that more prior control actions were remembered in the Writing condition than in the No-Writing condition. However, a repeated-measures design was used such that the No-Writing condition always preceded the Writing condition. Therefore, it is not clear whether memory differences were due to practice or condition effects. Furthermore, the participants used in this study were not air traffic controllers and therefore, the results provide only minimal support for the interactionist position of Hopkin (1988) and others. No differences in memory for previously performed actions were found when the same basic experiment was conducted using actual controllers (Zingale, Gromelski, Ahmed, & Stein, 1993).

Further evidence contradicting the hypotheses of the interactionist view has been provided by studies that demonstrate a lack of detrimental effects on performance, workload, or cognitive processing when controllers were limited in the amount of interaction they had with the FPS or when the FPS were completely removed (Albright, Truitt, Barile, Vortac, & Manning, 1994; Vortac, Edwards, Fuller, & Manning, 1993). Vortac et al. observed controllers (FAA Academy Instructors) under both normal and restricted FPS conditions. The FPS were posted and visible during the restricted condition, however, controllers could not physically manipulate or write on the FPS. Controller performance, including visual search and recall of flights and flight data, was not impaired by the restricted condition. In fact, controllers were more likely to remember to grant requests and did so sooner under the restricted condition. Vortac et al. concluded that by restricting interaction with the FPS, the ATC task was changed such that controllers were now able to assume a more strategic outlook. Restricting interaction with the FPS resulted in a reduction in workload, or at least a redistribution of workload in regard to the FPS, and more cognitive resources could be directed towards prospective activities.

The lack of evidence supporting a detrimental effect due to restricted interaction with the FPS suggests that whatever benefit controllers get from physical interaction with the FPS may be insufficient to result in a distinct memory or performance improvement. Because experts have better memory for task-specific information than novices, these results are in agreement with other research on expertise (Chase & Simon, 1973; Ericsson & Kintsch, 1995; Ericsson & Staszewski, 1989). Perhaps controllers generate enough information, and hence strong enough memory traces, through cognitive activity required to perform the ATC task that simply writing the information down or moving an FPS does not provide much additional benefit in terms of memory encoding. Furthermore, controllers may not be remembering the verbatim information as written down or seen on the plan view display (PVD). Instead, they may be remembering information regarding the gist of a situation or command (Gronlund, Ohrt, Dougherty, Perry, & Manning, 1998). Gronlund et al. also demonstrated that memory for aircraft information was not affected by the level of interaction with an aircraft, but was improved by the importance of the aircraft (i.e., likelihood of being in conflict). Therefore, the value of FPS interaction may not be in its benefits to memory encoding, but in its ability to aid in the retrieval or verification of verbatim information that could not otherwise be quickly retrieved from memory.

If the FPS serve only as aids to retrieval, then limiting interaction with the FPS, either partially or completely, should interfere with controller memory and performance to the extent that an alternate means of obtaining information is unavailable. If the functions served by the FPS are needed, then controllers should find ways to compensate for the lack of FPS interaction. The alternate means that controllers use to compensate, for example, by writing information on a notepad or retrieving more information from the computer system via a flight plan readout, tells us more about the role of the FPS in controller activity.

Albright, Truitt, Barile, Vortac, and Manning (1995) previously examined the role of the FPS by observing how controllers compensate for the absence of the FPS. They observed full performance level, en route controllers in a simulated, high altitude sector during both a normal condition and a condition in which all FPS were removed. By removing all FPS from the controllers and giving them a notepad on which to write anything they wished, Albright et al. were able to observe whether controllers could compensate for the lack of FPS and - if so - how they compensated. A subject matter expert (SME) evaluated controllers' performance and controllers provided subjective workload ratings after each scenario. Results showed no differences in performance or perceived workload when the FPS were absent. Controllers compensated for the lack of FPS by performing more flight plan readouts on the computer system. The flight plan readout provides the same basic information as does an FPS. Although this means of accessing flight information appeared to have slowed the time it took controllers to grant a request when FPS were absent, controllers spent significantly more time watching the PVD and wrote very little on the notepad. Given the results of Vortac et al.

(1993) and the way in which controllers compensated for the absence of FPS in the Albright et al. study, it is arguable that the primary function of the FPS is to provide ready access to flight information and, in terms of ultimate performance and memory, very little benefit is provided by writing on the FPS.

The results of Vortac et al. (1993) and Albright et al. (1995) suggest that, given the physical limitations of the DSR and the space allotted for the FPS, it may be practical to reduce the amount of writing on the FPS and the number of FPS that must be posted. Currently, en route controllers are required to post an FPS for an aircraft from the time radar contact and communications are established with that aircraft until the controller instructs the aircraft to switch radio frequencies and contact the next controller. These requirements result in at least one FPS being posted for each aircraft in a controller's airspace. Furthermore, many control actions must be recorded on the FPS as per the FAA Air Traffic Controller Handbook, 7110.65J (1995).

Reducing the requirements for FPS marking and posting by making them optional to the controller thus should not prevent the controller from achieving adequate performance while working within the current specifications of the DSR. Reducing the FPS requirements eliminates the redundant recording of information and potentially results in fewer FPS that the controller must post and search through in order to find important information. However, the evidence in support of reduced FPS marking and posting is somewhat limited. The Vortac et al. (1993) study was limited to individual controllers who were instructors at the FAA Academy. Although Albright et al. (1995) used field controllers, their results are also limited in that they only observed individual controller behavior in a single, high altitude sector. In order to provide further support for the viability of the reduction of FPS posting and marking, questions similar to those asked by Vortac et al. and Albright et al. must be addressed in a variety of both high and low altitude sectors. Furthermore, results must be generalizable beyond the individual controller, and the impact of reduced FPS activities must be assessed for the en route air traffic control team as well.

The current experiment was designed to answer two basic questions: First, does providing controllers with the option of posting and marking FPS result in significantly fewer FPS posted at any given time? Second, if the optional posting and marking of FPS does result in fewer FPS being posted, what, if any, are the effects on controller performance and workload? To answer these questions, participants were observed under both Normal and Optional FPS Marking/Posting conditions. During the Normal condition, controllers had full use of the FPS and they controlled traffic as they usually would. Under the Optional FPS condition, with some exceptions, controllers had to post and mark FPS only until radar contact and communications were established and accepted with an aircraft.

The procedures used in the Optional FPS condition were developed by The Strip Reduction Working Group, which met from November 4-6, 1997, in Washington, D.C. The sole purpose of the meeting was to identify ways to reduce en route controllers' use of FPS and FPS marking in anticipation of the DSR upgrade. The group was organized by the FAA Air Traffic Operations branch and included representatives from ATO-110, the Civil Aeromedical Institute, the National Air Traffic Control Association, the ATC

Supervisors Committee, and the University of Oklahoma. The varied composition of the Strip Reduction Working Group allowed for consideration of many possibilities for the reduction of FPS activity. The group discussed and recommended changes to the current FPS procedure as possible ways to reduce FPS activity. The result of the group's effort was an alternative FPS procedure that is shown in Appendix A. An important point of the revised FPS procedure is that although it allows for a reduction in the posting and marking of FPS, such a reduction is optional and would only be used if the controller responsible for a sector decided to do so. Before adjourning, the Strip Reduction Working Group decided that an empirical study would be appropriate to test the revised FPS procedure.

The experiment was conducted at two Air Route Traffic Control Centers (ARTCC), Cleveland (ZOB) and Jacksonville (ZJX)[1]. Data on performance and workload were collected from individual controllers and controller teams operating in either high or low altitude sectors.

Method

Participants

A total of 48 full performance level (FPL) controllers volunteered to participate in the experiment (ZOB=24, ZJX=24). Just prior to the experiment, each controller read and signed an informed consent statement and then completed a biographical questionnaire (shown in Appendix B). Mean responses to the biographical questionnaire are shown in Table 1. Controllers participated either individually or as part of a two-

[1] Additional data focusing on individual controllers in low altitude sectors were collected at Boston (ZBW). Because there was no effect of the experimental manipulation at ZBW, and because these data

person team. Controller teams are generally made of two people; one person (the R-side) is primarily involved with activities associated with the radar screen while the second person (the D-side) is mainly involved with posting and marking the FPS. The D-side also handles the transfer of information to, and the recording of information from, other sectors and ARTCCs.

Table 1. Means and standard deviations for biographical data by ARTCC.

Biographical Info	**ZOB**	**ZJX**
Number of ARTCCs worked	1.17 (0.38)	1.17 (0.38)
Years in current area	10.37 (7.01)	10.75 (7.04)
Years at current ARTCC	11.28 (6.73)	11.86 (6.56)
Years as FPL	8.94 (6.91)	11.15 (8.07)
Years as Controller	13.66 (6.49)	15.30 (6.84)
Time since recertification (yrs.)	4.91 (4.60)	7.89 (7.77)
Age	38.46 (6.38)	39.13 (6.65)

Materials and Equipment

Scenarios

For each ARTCC, two scenarios were developed for each sector (high or low altitude) and staffing condition (individual or team) for a total of eight scenarios per ARTCC. All scenarios were selected from training scenarios that had been developed previously by the training department of each ARTCC. Once selected, scenarios were altered if necessary to meet the complexity requirements of the experiment or to insure the occurrence of particular events of interest. Sectors used at ZOB were Hudson (high altitude) and Lichtfield (low altitude). Sectors used at ZJX were Brewton (high altitude) and Florence (low altitude). General descriptions of these sectors are provided in

focused on individuals in low altitude sectors only, these data are not presented here.

Appendix C. All scenarios were designed to be at least 30 minutes in length. Scenarios for individual controllers had a complexity rating of 70% and scenarios for controller teams had a complexity rating of 100%. Complexity ratings were calculated by the ARTCC's training specialists and were based on the number and type of events that occurred during each scenario.

Within the typical occurrences of each scenario, a strip-critical event was identified in order to give the SME an opportunity to evaluate how each event was handled, especially when the related FPS had already been removed from the board. Each strip-critical event required the controller to make use of an FPS and was yoked to the scenario in which it occurred. Strip-critical events included, 1) providing holding instructions, 2) a pilot requesting a route change, 3) an aircraft flying at wrong altitude for direction of flight, and 4) an aircraft requiring special handling such as Air Force One.

Dynamic simulator (DYSIM) training facility

The DYSIM is a high fidelity simulation facility that closely resembles the real en route ATC environment. Workstations are fully functional and landline communications are provided. Flight plans for each aircraft in a scenario are pre-programmed, but controllers or simulation pilots may change any flight plan during the simulation. Training specialists execute ATC instructions to simulate the roles of pilots and other controllers during the simulation.

Workload Assessment Keypad (WAK)

The WAK is a computer-controlled, on-line subjective measure of workload. The WAK is well adapted for use in the field because it is relatively small, portable, and can collect ratings from as many as four participants simultaneously. Based on the work of Stein (1985), the WAK obtains a workload rating by auditorily and visually prompting the participant. At an adjustable interval, the WAK emits a high-pitched tone and its seven, numbered buttons illuminate. The participant then makes a rating by pressing one of the buttons within a specified amount of time. The WAK records each rating as well as elapsed time from prompt to response. The WAK instructions are shown in Appendix D.

Task Load Index (TLX)

A modified version of the NASA TLX (Hart & Staveland, 1988) was used to collect subjective ratings of taskload. The TLX contains six separate scales to assess mental demand, physical demand, temporal demand, effort, frustration, and performance. Each scale was represented as a 100mm line, anchored from low to high.

The TLX rating form, instructions, and scale descriptors are shown in Appendix E. Controllers placed an "X" on each scale after an experimenter described what the scale was intended to measure.

Procedure

The procedure was the same for both ARTCCs and is summarized in Table 2. All data were collected in the DYSIM training facility of the respective ARTCC. Once controllers arrived at the DYSIM, they completed an informed consent statement and biographical questionnaire. An experimenter then provided instructions on how to use the WAK.

Once controllers understood the WAK instructions, data collection began with the first of two scenarios. The first scenario was always the Normal condition in which participants were asked to control traffic as they normally would. Although the order of conditions (Normal vs. Optional FPS) was not counterbalanced, the two scenarios corresponding to group (individual or team and high or low altitude sector) were counterbalanced, such that each scenario appeared in each condition an equal number of times. Complete counterbalancing for two factors (Scenario and Condition) with two levels each was not possible because there were only four participants from each ARTCC in each group. Participants were given a notepad to record anything they wished. The scenario began with the SME providing a position relief briefing and then the participant took full control of the scenario for 30 minutes.

During the scenario, the WAK prompted the participant every 5 minutes for a workload rating (where 1 = very low workload and 7 = very high workload). The SME used a behavioral event checklist (shown in Appendix F) to record the occurrence of specific events related to controller performance. The SME also observed one strip-critical event (for example, pilot requests route change) and noted if the participant effectively handled that event. For controller teams, communication effectiveness was evaluated by the SME using the form shown in Appendix G.

Two experimenters used microcassette recorders to archive activity relevant to the plan view display (PVD), computer readout display (CRD), and flight progress strips (FPS). In addition, experimenters recorded the time at which each activity occurred. One experimenter recorded the type of actions performed by the controller involving the PVD and CRD and the time these actions were performed. These actions included request for information, use of a route display[2], J-ring[3], /0[4], and flight plan readout (FPR).[5] The second experimenter recorded the callsigns of FPS and the times each was posted and removed. The second experimenter also recorded D-side activities regarding the PVD and D-side CRD when controller teams were being observed. At the end of the

[2] A route display is indicated on the PVD as a line drawn from a selected aircraft's current position to that aircraft's position at some designated time, from 1 to 99 minutes, into the future. The route display is based on flight plan information that is currently stored in the computer. The route display is automatically shown whenever a route amendment is made to an aircraft's flight plan via computer entry by the controller.

[3] A J-ring (also referred to as a distance reference indicator) is displayed on the PVD as an approximate circle that can be placed around selected aircraft for means of determining horizontal separation. The J-ring is an approximate circle and has a parameter-defined number of sides. The radius of the J-ring is typically 5 miles but this parameter can also be set by each ARTCC.

[4] A "slant zero", /0, is an action performed by the controller via keyboard entry to shorten the length of a leader line (the line connecting an aircraft position symbol with a datablock as displayed on the PVD). Many, but not all, controllers use the /0 entry as an indicator, or reminder, that they have instructed an aircraft to switch radio frequencies. Because all aircraft are instructed to switch frequencies when departing one sector of airspace and entering another, controllers who use the /0 entry performed it on virtually all aircraft departing their sector.

[5] The controller can display a FPR on the CRD via a Quick Action Key and keyboard entries. The FPR displays much of the critical information that is displayed on the FPS including, callsign, computer identification number, beacon code, assigned altitude, and route of flight information. The FPR is

first scenario, the participant used the computer-based (CRD) checklist to provide a position relief briefing to the SME. Once completed, the SME evaluated the quality of the briefing using the form shown in Appendix H. The participant then completed the TLX followed by a 15-minute break.

Participants returned to the DYSIM after the break. Before starting the Optional FPS condition an experimenter reminded the participants how to use the WAK. A representative of the National Air Traffic Controllers Association (NATCA) then summarized the proposed strip marking and posting procedure. Under the Optional FPS condition, controllers were instructed by the NATCA representative that they were to post and mark FPS only until radar contact and communications were established and accepted with an aircraft. After that, strip posting and marking became optional. Participants placed a check mark in field 21-24 of a strip to indicate that optional posting and marking could be used for that strip. However, optional marking only applied to information that was recorded elsewhere, for example, by computer entry or voice recordings. The NATCA representative also instructed participants that information that was not redundant had to be recorded on the FPS and that an FPS could be removed from the board if it was no longer needed. Participants were also instructed that an FPS was required to be posted and marked in special situations which included, 1) radar contact would be lost, 2) an aircraft was transitioning from radar to non-radar, 3) special handling was required, 4) non-radar flight, 5) an aircraft transitioning from auto to non-auto mode, and 6) holding instructions issued. The full text of the Optional

---

displayed indefinitely until another action resulting in a CRD message overwrites it.

FPS usage procedure is shown in Appendix A. Finally, participants were encouraged (but not required) by the NATCA representative to follow the optional FPS posting and marking procedures as best they could so that an adequate test of the procedure could be conducted. Experimenters then provided each controller with a Procedures Summary Sheet and reviewed each item on the sheet with the controller. The Summary Sheet is shown in Appendix I. As in the Normal condition, a notepad was provided for the controller to record anything he or she wished.

Once the participant indicated that he or she understood the instructions, the Optional FPS condition began with a position relief briefing from the SME after which the participant took full control of the scenario for 30 minutes. As before, the SME and two experimenters observed the participant's activity, and controllers were prompted by the WAK every 5 minutes for a workload rating. The participant used the computer-based (CRD) checklist to provide a position relief briefing at the end of the scenario and then completed the TLX a second time. Finally, controllers were debriefed, thanked, and released.

Table 2. Summary of steps in experimental procedure.

1	Participant, experimenters, SME, & simulation pilots meet in DYSIM
2	Participant completes informed consent statement & biographical questionnaire
3	Normal condition begins – Participant receives WAK instructions
4	Participant instructed by experimenter to control traffic normally
5	Scenario is started & participant receives position relief briefing from SME
6	Data collection for 30 minutes
	Two experimenters record activity regarding PVD, CRD, and FPS
	SME records strip-critical & behavioral events & communication effectiveness
	Participant controls air traffic & makes WAK rating every 5 minutes
7	Participant provides position relief briefing to SME
	SME evaluates completeness of briefing, noting missed items
8	Participant completes TLX
9	15-minute break
10	Participant, experimenters, SME, & simulation pilots return to DYSIM
11	Optional FPS condition begins – Participant receives WAK instructions
12	Participant receives instructions on Optional FPS procedure from NATCA
13	Participant reviews Optional FPS procedures summary sheet
14	Repeat steps 6-8.
15	Participant completes post-experimental questionnaire
16	Participant receives debriefing from experimenter

Results

Data from ZOB and ZJX were combined, yielding a total of 16 individuals (8 per sector type) and 16 teams (8 per sector type) for the analysis. Results of individuals and teams will be reported separately for each sector type. Evaluative comparisons will not be made between ARTCCs, individuals and teams, or high and low altitude sectors because the objective of this experiment was to evaluate the Optional FPS procedure, not to evaluate a particular facility. It should be noted that scenarios were used an equal number of times in each condition. Therefore, any differences between the Normal and Optional FPS conditions were not due to differences in scenarios. Statistical values are only reported for tests that were significant at a level of $\alpha < .05$.

Proportion of total FPS posted

It would be difficult to assess the reduced posting and marking procedure without some willingness on the participant's part to try the new procedure. An examination of the proportion of total FPS that could have been posted during the scenarios provides information regarding whether or not participants were willing and able to follow instructions and if in fact the Optional FPS procedure resulted in fewer FPS being posted over time. Figures 1 through 4 show the mean proportion of total FPS posted by condition and scenario time for individuals and controller teams in both high and low altitude sectors.

Each of the four datasets were analyzed using a 2 (Normal vs. Optional FPS) X 30 (1-min intervals) repeated measures analysis of variance (ANOVA). When analyzing the proportion of FPS posted, individuals in the low altitude sectors posted proportionally fewer FPS in the Optional FPS condition ($M = 0.49$, $SD = 0.15$) than in the Normal condition ($M = 0.64$, $SD = 0.16$), and posted fewer FPS over time, $F(1, 7) = 15.50$ and $F(29, 203) = 14.13$, respectively. There was a significant Condition X Time interaction, $F(29, 203) = 4.71$. Individuals in the low altitude sector were able to use the Optional FPS procedure to reduce the proportion of FPS posted during the first few minutes of the scenario and then retained that reduction.

Individuals in high altitude sectors posted proportionally fewer FPS in the Optional FPS condition ($M = 0.50$, $SD = 0.18$) than in the Normal condition ($M = 0.58$, $SD = 0.19$), $F(1, 7) = 23.92$, and the number of FPS posted declined over time, $F(29, 203) = 3.15$. The Condition X Time interaction was not significant indicating that the proportional reduction was relatively immediate and was retained throughout the scenario. Individuals in high altitude sectors were able to implement the Optional FPS procedure in that they were able to post proportionally fewer strips in the Optional FPS condition than in the Normal condition.

Controller teams in low altitude sectors posted proportionally fewer FPS in the Optional FPS condition ($\underline{M}$ = 0.48, $\underline{SD}$ = 0.16) as compared to the Normal condition ($\underline{M}$ = 0.62, $\underline{SD}$ = 0.17), $\underline{F}(1, 7) = 8.36$, and also posted fewer FPS over time, $\underline{F}(29, 203) = 14.72$. The Condition X Time interaction was also significant, $\underline{F}(29, 203) = 6.86$. Therefore, teams in low altitude sectors used the Optional FPS procedure to reduce the proportion of FPS that were posted early in the scenario relative to the Normal condition. Although this reduction took some time, the reduction was maintained after the first few minutes of the scenario.

Controller teams in high altitude sectors posted proportionally fewer FPS in the Optional FPS condition ($\underline{M}$ = 0.45, $\underline{SD}$ = 0.16) than in the Normal condition ($\underline{M}$ = 0.62, $\underline{SD}$ = 0.17), and posted fewer FPS over time, $\underline{F}(1, 7) = 81.19$ and $\underline{F}(29, 203) = 39.09$, respectively. There was a significant Condition X Time interaction, $\underline{F}(29, 203) = 8.36$. Similar to teams in low altitude sectors, teams in high altitude sectors also were able to use the Optional FPS procedure to reduce the proportion of FPS posted during the first few minutes of the scenario and then retained that reduction. Controller teams in both high and low altitude sectors removed FPS at a greater proportional rate during the Optional FPS condition. Overall, participants were able to post proportionally fewer FPS during the Optional FPS condition as compared to the Normal condition.

A significant main effect of Time was found for all four groups of participants. That is, proportionally fewer FPS were posted over time regardless of condition (Normal or Optional FPS). This effect is best explained by the fact that the scenarios tended to slow down somewhat as they progressed. Because the scenarios were designed to be 30 minutes in length, new aircraft were not being generated towards the end of the scenarios and hence, the number of FPS that were posted tended to decrease over time. To control for the effect of Time, difference scores were calculated for each group of participants by subtracting the proportion of FPS posted in the Optional FPS condition from the proportion of FPS posted in the Normal condition for each minute of the scenario. These data are shown in Figures 5 through 8.

Overall, the difference scores show that the reduction in the proportion of FPS posted in the Optional FPS condition as compared to the Normal condition occurred relatively early in each scenario. The difference in the proportion of FPS posted was then maintained throughout the scenarios. Individuals in the high altitude sectors showed the smallest difference in the proportion of FPS posted with a reduction of 5% to 15% fewer FPS posted in the Optional FPS condition as compared to the Normal condition. The greatest reduction in the proportion of FPS posted was realized by the controller teams in the high altitude sectors who posted at times a difference greater than 20% fewer FPS in the Optional FPS condition. Although these proportional differences may seem small, especially when considering only a single sector of airspace, the differences have much greater consequences when one considers the overall impact of a procedure similar to the one used in the Optional FPS condition when implemented within the entire en route ATC system.

Number of marks per flight progress strip

In addition to observing how many FPS were posted, the average number of marks made on each strip were also counted. Again, this insured that participants were in fact using the optional marking procedure as instructed. Participants did in fact make about 1 less mark per strip in the Optional FPS condition than in the Normal condition. Data are shown in Figures 9 through 12.

Data were analyzed using a dependent t-test for each data set. Figure 9 shows the mean number of marks per strip for individuals in low altitude sectors. Significantly fewer marks were made in the Optional FPS condition (M = 4.19, SD = 0.20) than in the Normal condition (M = 4.8, SD = 0.15), $t(1, 7) = 2.64$. Individuals in high altitude sectors also made significantly fewer marks per FPS in the Optional FPS condition (M = 2.52 , SD = 0.10) than in the Normal condition (M = 3.65, SD = 0.13), $t(1, 7) = 3.05$. Data are shown in Figure 10. Individuals in both low and high altitude sectors made significantly fewer marks on the FPS in the Optional FPS condition as compared to the Normal condition.

Figure 11 shows that controller teams in low altitude sectors made significantly fewer marks in the Optional FPS condition (M = 2.39, SD = 0.11) as compared to the Normal condition (M = 3.74, SD = 0.49), $t(1, 7) = 3.19$. Controller teams in high altitude sectors made a comparable number of marks per FPS in the Optional FPS condition (M = 2.93, SD = 0.07) and the Normal condition (M = 3.35, SD = 0.19). Data are shown in Figure 12. With the exception of teams in high altitude sectors, participants made fewer marks on the FPS under the Optional FPS condition. This result supports the fact that participants were implementing the experimental procedure.

On-line workload ratings

It was expected that on-line rating of workload using the WAK would be lower during the Optional FPS condition due to the reduced board management responsibilities. Alternatively, the introduction of a new procedure could produce more board management duties and hence, more workload. The WAK data are shown in Figures 13 through 18 which present data from both high and low altitude sectors provided by individuals and team R-sides and D-sides. Missing data due to a participant not responding to a single WAK prompt were replaced by the appropriate group mean. Replacement by group means artificially reduces variance.[6] All controllers were included in the analysis even if they failed to respond to more than one WAK prompt. Each of the 6 datasets was analyzed separately using a 2 (Normal vs. Optional FPS) X 6 (5-min intervals) repeated measures ANOVA.

The results, as shown in Figure 13, indicate that individual controllers in low altitude sectors rated workload as being significantly lower in the Optional FPS condition, $F(1, 7) = 23.69$, and as increasing over time, $F(5, 35) = 19.00$. Individual controllers in high altitude sectors also rated the Optional FPS as having lower workload, but this difference was not significant. This result coincides with the finding of no difference in the number of FPS that were posted by the same individuals.

[6] The replacement of missing data in subjective workload measurement remains an unresolved procedural issue. It is preferable to replace missing WAK data with group means rather than with a maximum workload rating because it was likely that a failure to respond was due to inadequate auditory and visual prompts rather than extremely high workload.

Individuals in high altitude sectors also rated workload as increasing significantly over time, $\underline{F}(5, 35) = 6.47$. These data are shown in Figure 14.

Separate WAK ratings were obtained for both the R-side and D-side controllers when controller teams were being observed. R-side controllers in low altitude sectors perceived workload as being less on average in the Optional FPS condition, but not significantly so. They did perceive workload to be increasing over time, $\underline{F}(5, 35) = 6.87$, as shown in Figure 15. Likewise, R-side controllers in high altitude sectors, shown in Figure 16, rated the Normal and Optional FPS conditions as being similar in workload yet increasing significantly over time, $\underline{F}(5, 35) = 10.52$. D-side controllers in both low and high altitude sectors, shown in Figures 17 and 18, rated workload as being comparable under the two FPS procedures. D-side controllers in high altitude sectors rated workload as increasing significantly over time, $\underline{F}(5, 35) = 6.29$.

Overall, participants tended to rate workload as low to moderate. Workload fluctuated and was generally perceived by participants, with the exception of low-altitude D-sides, as changing throughout the scenario. Participants judged workload in the Optional FPS condition as being comparable and occasionally lower than in the Normal condition.

It is not clear why participants judged workload as increasing over time. Thirty minutes is not an unusually long time for a controller to set at a position so fatigue is an unlikely explanation. More likely, differences in subjective workload ratings may be attributed to participant characteristics in which the actual rating scale that is used may have changed as the scenarios progressed. In other words, workload may not have actually changed over time, but the way that participants rated workload did change. A level of workload that was rated as being low at the start of a scenario may have been rated higher later in the scenario even though the same level of workload existed. Item response theory is a way in which participant and scenario characteristics may be examined in more detail (see Appendix J).

Time to respond to WAK

The WAK ratings were completely subjective and hence, a more objective measure of workload may be more sensitive to changes in taskload. Because participants were instructed to respond to the WAK prompt without letting it interfere with the primary task of controlling traffic, the time it took participants to respond to the WAK may be considered as a secondary workload probe and a more objective measure of workload. Time to respond to the WAK prompt was analyzed to determine if this measure would provide a more sensitive estimate of workload. Like the WAK ratings, time to respond to the WAK (WAK RT) was analyzed for each of the 6 datasets. Each dataset was analyzed separately using a 2 (Normal vs. Optional FPS) X 6 (5-min intervals) repeated measures ANOVA. Results are shown in Figures 19 through 24.

Significant differences were obtained only for R-side controllers in high altitude sectors. Significant main effects for both Condition, $\underline{F}(5, 35) = 10.11$, and Time $\underline{F}(5, 35) = 2.96$ were found. The Condition X Time interaction was not significant. Therefore, the time to respond to the WAK showed that participants serving as R-side controllers in high altitude sectors had a greater workload in the Normal condition as compared to the Optional FPS condition and that the workload was fluctuating over time. Even though the time to respond the WAK may be a more objective estimate of workload, this measure did not indicate any increase in workload for the Optional FPS condition. In fact, workload in the Optional FPS condition can still be said to be equivalent, and sometimes less, than in the Normal condition. Furthermore, the increases in workload over time demonstrated by the subjective ratings were not evident (with the exception of R-side controllers in high altitude sectors) when time to respond to the workload ratings was considered.

Post-scenario TLX ratings

Participants provided another subjective rating of workload after each scenario by completing the TLX. The TLX ratings were analyzed separately for each group of participants (Individual, R-side, and D-side in both high and low altitude sectors) using a 2 (Optional FPS vs. Normal) X 6 (TLX item) multivariate analysis of variance (MANOVA). None of the omnibus multivariate analyses were significant and no further analyses were pursued. Results are shown in Figures 25 through 30. Although results were not significant, some small differences are visible and the results were graphed for this purpose. The TLX results only suggest weak relationships in the data and were not supported by statistical analyses. Overall, the TLX ratings were similar for the Optional FPS and Normal conditions. The TLX ratings were also examined using item response theory as described in Appendix J.

Compensatory behaviors

Participants could have written on the notepad to compensate for marking the FPS less often. However, participants wrote on the notepad infrequently. Participants made a comparable number of marks on the notepads for the Normal and Optional FPS conditions. An average of 0.5 marks were made on the notepad in the Normal condition and an average of 1.1 marks were made on the notepad in the Optional FPS condition. The small number of notes written suggest that either participants did not think writing information down was necessary or that using the notepad would have required too much work and so it was not used very often. Notes often referred to information that was not normally required but was needed to operate within the DYSIM, for example, the sector number receiving a hand-off.

In addition to the notepad, participants could compensate for reduced FPS posting and marking requirements by utilizing available computer functions such as the flight plan readout (FPR) or route display. The number of compensatory actions were analyzed using a 2 (Normal vs. Optional FPS) X 4 (FPR, route display, J-ring, slant-0) within-subjects MANOVA. No significant differences in the number of compensatory actions between the Normal and Optional FPS conditions were found. Apparently, participants did not make significant changes in their behavior to compensate for the reduction in FPS activity as evidenced by their lack of writing on the notepads and use of computer-based functions. Although not statistically significant, participants tended to use slightly more FPRs in the Optional FPS condition than in the Normal condition. Compensatory behavior data is shown in Tables 3-5.

Table 3. Compensatory behavior: Means (standard deviations) for individual controllers by sector type and condition.

Sector	Condition	FPR	Route Display	J-ring	/0
Low	Normal	2.75	2.00	1.50	9.13
		(3.58)	(1.31)	(2.07)	(4.49)
	Optional FPS	4.75	2.75	1.63	9.63
		(8.03)	(1.83)	(1.60)	(3.85)
High	Normal	2.75	4.50	0.63	10.38
		(4.40)	(1.60)	(0.92)	(9.24)
	Optional FPS	1.75	5.75	0.00	10.00
		(1.58)	(3.06)	(0.00)	(7.31)

Table 4. Compensatory behavior: Means (standard deviations) for team R-side controllers by sector type and condition.

Sector	Condition	Point FPS	Point PVD	Request Info	FPR	Route Display	J-ring	/0
Low	Normal	0.25	2.63	2.38	1.88	1.50	1.63	9.88
		(0.46)	(1.85)	(1.30)	(2.30)	(1.07)	(1.60)	(4.97)
	Optional FPS	0.13	1.50	2.38	2.88	1.50	1.50	8.25
		(0.35)	(1.31)	(2.77)	(3.09)	(1.41)	(1.51)	(4.46)
High	Normal	1.75	2.25	1.00	2.00	2.75	1.13	13.50
		(2.19)	(1.91)	(1.20)	(1.51)	(1.28)	(1.89)	(5.04)
	Optional FPS	0.13	2.00	1.25	4.75	3.75	1.33	13.50
		(0.35)	(1.69)	(1.04)	(4.43)	(2.49)	(2.26)	(6.02)

Table 5. Compensatory behavior: Means (standard deviations) for team D-side controllers by sector type and condition.

Sector	Condition	Point FPS	Point PVD	Request Info	FPR	Route Readout	/0
Low	Normal	2.75	2.00	0.38	0.50	1.25	0.00
		(1.83)	(1.41)	(1.06)	(0.76)	(1.16)	(0.00)
	Optional FPS	1.75	2.50	0.25	1.00	0.88	0.13
		(1.67)	(2.98)	(0.46)	(1.31)	(1.36)	(0.35)
High	Normal	3.25	1.63	0.50	1.38	3.25	0.25
		(2.31)	(0.92)	(1.07)	(0.74)	(3.11)	(0.71)
	Optional FPS	3.63	3.75	1.13	3.38	2.63	0.38
		(2.67)	(2.31)	(1.25)	(2.92)	(3.46)	(1.06)

<u>Subject Matter Expert observations</u>

An SME for each sector used the Behavioral and Event Checklist to record the occurrence of 11 events including operational errors, operational deviations, missed handoff, violation of a Letter of Agreement or other directive, missed readback error, failure to grant request, failure to direct aircraft to switch frequency, cause unnecessary delay, inappropriate request of information, computer entry error, and failure to complete proper coordination. Subject matter experts selected these particular events as types of events that may have a negative impact upon operations, especially during the Optional FPS condition. Results are shown in Figures 31 through 34. Events on the checklist seldom occurred, but when one did it was just as likely to occur in either experimental condition. One operational error[7] did occur for a controller team in a high altitude sector but the error was not related to FPS activity. Overall, the Behavioral and Events Checklist did not detect any significant differences between the Normal and Optional FPS conditions for number and type of events that occurred.

Like behavioral events, it was possible that the Optional FPS condition might impact team communication effectiveness and the adequacy of position relief briefings. On the contrary, these measures showed no differences between conditions. Team communication, as noted by our SMEs, was not adversely impacted. Only one negative comment was made by an SME regarding team communication. However, this comment occurred during the Normal condition and was related to participants having to repeat an action unnecessarily. Position relief briefings did not suffer either. The SMEs did not note any deficiencies regarding position relief briefings in either the Normal or Optional FPS condition.

[7] Aircraft must be separated by 5 nautical miles (NM) laterally or 1,000 feet vertically when flying below 29,000 feet mean sea level (MSL). When flying at, or above, 29,000 feet MSL aircraft must be separated by 5 NM laterally or 2,000 feet vertically. An operational error occurs when two or more aircraft violate these separation standards.

Post-scenario questionnaire

Data from the post-scenario questionnaire are shown in Table 6. The concerns regarding realism typically referred to situations that were imposed by the idiosyncrasies of the DYSIM facility. Although each pair of scenarios used were constructed and evaluated by SMEs to be similar in complexity, some participants perceived the scenarios they controlled as being similar in complexity, while others did not. Likewise, some participants mentioned that either their counterpart (R-side or D-side) or a simulation pilot had done something out of the ordinary during the experiment. None of the participants indicated that the WAK measure interfered with their ability to control traffic. Finally, many of the participants reported that they preferred the optional posting and marking procedures.

Table 6. Percentage of participants by group responding "yes" to items on the post-scenario questionnaire.

	Individual		R-side		D-side	
	Low	High	Low	High	Low	High
Did you notice anything unusual or unrealistic about either of the scenarios?	75.0%	25.0%	87.5%	25.0%	37.5%	37.5%
Were the scenarios similar in complexity?	50.0%	100.0%	87.5%	75.0%	50.0%	37.5%
Did the pilots or team member do anything strange?	25.0%	0.0%	0.0%	37.5%	25.0%	12.5%
Did responding to the workload rating hinder your ability to control traffic?	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Did you prefer the optional strip marking procedure?	75.0%	100.0%	75.0%	75.0%	87.5%	62.5%
Did you prefer the optional strip posting procedure?	87.5%	75.0%	62.5%	75.0%	87.5%	62.5%

Conclusions

Overall, participants at ZOB and ZJX posted fewer FPS and made fewer marks on those that were posted during the Optional FPS condition. Even though FPS activity was reduced in the Optional FPS condition, no detrimental effects in performance, workload, position relief briefings, or team communication were observed. It is important to note that participants performed similarly in both experimental conditions despite never having practiced using the optional FPS posting and marking procedure prior to the experiment. According to the post-experimental questionnaire, most participants preferred the optional FPS marking and posting procedures they used during the Optional FPS condition. Participants did not compensate for the lack of FPS by using other tools, such as the FPR or notepad, to obtain or remember information that would have otherwise been present on an FPS. No detrimental effects of the Optional FPS condition were detected. Therefore, the results of the present experiment suggest that the Optional FPS condition provided a viable procedure by which FPS activity could be reduced.

Although the data collected in the present study provide encouraging support for the use of an optional FPS posting and marking procedure, some issues remain. Participants were engaged in the scenarios for a relatively short period of time. Participants generally had a high degree of vigilance given that experimenters and an SME were observing them. It is possible that given more time, general patterns of behavior may change as participants become more relaxed. Additionally, given more time to familiarize themselves with the Optional FPS procedure, it is likely that participants may develop and adopt behaviors unlike those observed in the experiment. In particular, it is possible that with longer periods of time to use the Optional FPS procedure, similar events may be more likely to interfere, that is, be confused with one another (McGeoch, 1942) and some means of compensation, such as cues for memory retrieval, may become more important. Controllers who interact with the same pilots and aircraft flying the same routes every day may be especially susceptible to memory interference because of the high degree of episodic similarity in temporally distinctive events.

Another question for future research regards how responsibilities are passed from one controller to the next during and immediately after a position relief briefing. If the reduced requirements for FPS management are adopted as optional and FPS use is at the discretion of the controller currently in charge of a sector, then how would responsibility be passed given that the controller relieving the position wishes to use a procedure different from the previous controller? The obvious problem occurs when the current controller is using the reduced posting and marking requirements and is then relieved by a controller who wishes to use and mark all available FPS. The question of how to effectively transition from one controller to another controller who wishes to use a different procedure remains to be answered.

Even though there are still a number of questions to be answered regarding how controllers will use the FPS in the future, the current study has made a significant contribution to the field of ATC research. Namely, the current study has shown that the Optional FPS procedure may be a viable means to reduce the controllers' reliance on the FPS. Such a reduction in the use and reliance on the FPS would bode well for the impending transition from the old M-1 console to the newer DSR console in that controllers would be able to make the transition without having to dramatically change the way that they use the FPS. Moreover, reducing the reliance on the FPS now will better prepare controllers for the likely event in which the paper FPS become partially or fully automated in an electronic format.

References

Albright, C. A., Truitt, T. R., Barile, A. L, Vortac, O. U., & Manning, C. A. (1994). Controlling traffic without flight progress strips: Compensation, workload, performance, and opinion. *Air Traffic Control Quarterly, 2,* 229-248.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores* (chapters 17-20). Reading, MA: Addison-Wesley.

Chase, W. G., & Simon, H. A. (1973). *Perception in chess.* Cognitive Psychology, *4,* 55-81.

Federal Aviation Administration (1995). *Air traffic control handbook* (7110.65J). Washington, DC: Author.

Ericsson, K. A., & Kintsch, W. (1995). *Long-term working memory.* Psychological Review, *102(2),* 211-245.

Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr and D. Kotovsksy (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp.235-267). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Federal Aviation Administration (1995). *Air traffic control handbook* (7110.65J). Washington DC: Author.

Gromelski, S., Davidson, L., & Stein, E. S. (1992). *Controller memory enhancement: Field facility concepts and techniques.* (DOT/FAA/CT-TN92/7). Atlantic City, NJ: U. S. Department of Transportation, Federal Aviation Administration.

Gronlund, S. D., Ohrt, D. D., Dougherty, M. R. P., Perry, J. L., & Manning, C. A. (1998). *Role of memory in air traffic control.* Journal of Experimental Psychology: Applied, *4,* 263-280.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: North-Holland.

Hopkin, V. D. (1988). Air traffic control. In E. L. Wiener and D. C. Nagel (Eds.), *Human factors in aviation* (pp. 639-663). San Diego: Academic Press.

Hopkin, V. D. (1995). Situational awareness in air traffic control. In R. D. Gilson, D. J. Garland, & J. M. Koonce (Eds.), *Situational awareness in complex systems: Proceedings of a CAHFA conference,* 171-178. Daytona Beach: Embry Riddle Aeronautical University Press.

Lord, F. M. (1952). *A theory of test scores* (Psychometric Monograph No. 7). Iowa City, IA: Psychometric Society.

McGeoch, J. A. (1942). *The psychology of human learning.* New York: Longmans, Green.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Iowa City, IA: Psychometric Society.

Slamecka, N. J., & Graf, P. (1988). *The generation effect: Delineation of a phenomenon.* Journal of Experimental Psychology: Human Learning & Memory, *4(6), 592-604.*

Stein, E. S. (1985). *Air traffic controller workload: An examination of workload probe.* (DOT/FAA/CT-TN84/24). Atlantic City, NJ: U. S. Department of Transportation, Federal Aviation Administration.

Stein, E. S. (1991). *Air traffic controller memory - A field survey.* (DOT/FAA/CT-TN90/60). Atlantic City, NJ: U. S. Department of Transportation, Federal Aviation Administration.

Stein, E. S., & Bailey, J. (1994). *The controller memory guide.* (DOT/FAA/CT-TN94/28). Atlantic City, NJ: U. S. Department of Transportation, Federal Aviation Administration.

Vortac, O. U., Edwards, M. B., Fuller, D. K. & Manning, C. A. (1993). *Automation and cognition in air traffic control: An empirical investigation.* Applied Cognitive Psychology, 7, 631-651.

Zingale, C., Gromelski, S., Ahmed, S. B., & Stein, E. S. (1993). *Influence of individual experience and flight strips on air traffic controller memory/situational awareness.* (DOT/FAA/CT-TN93/31). Atlantic City, NJ: U. S. Department of Transportation, Federal Aviation Administration.

Zingale, C., Gromelski, S., & Stein, E. S. (1992). *Preliminary studies of planning and flight strip use as air traffic controller memory aids.* (DOT/FAA/CT-TN92/22). Atlantic City, NJ: U. S. Department of Transportation, Federal Aviation Administration.

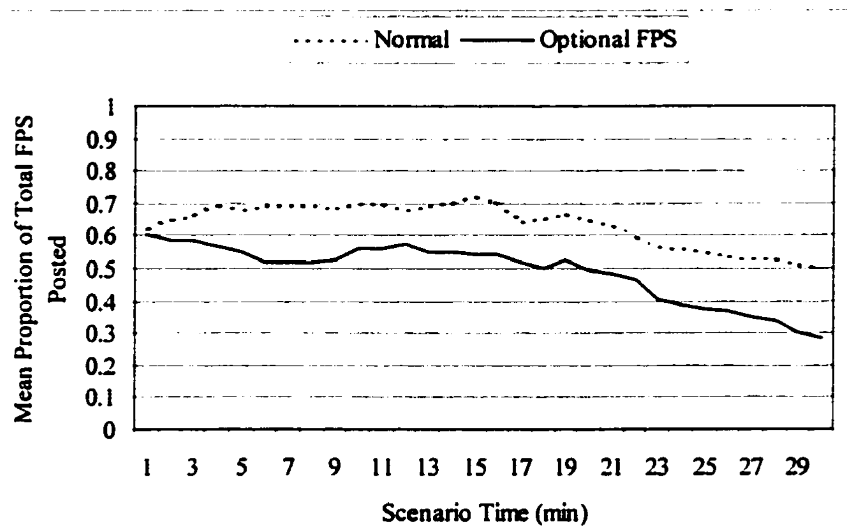


Figure 1. Mean proportion of total FPS posted by condition and scenario time for Individual controllers in Low Altitude sectors.

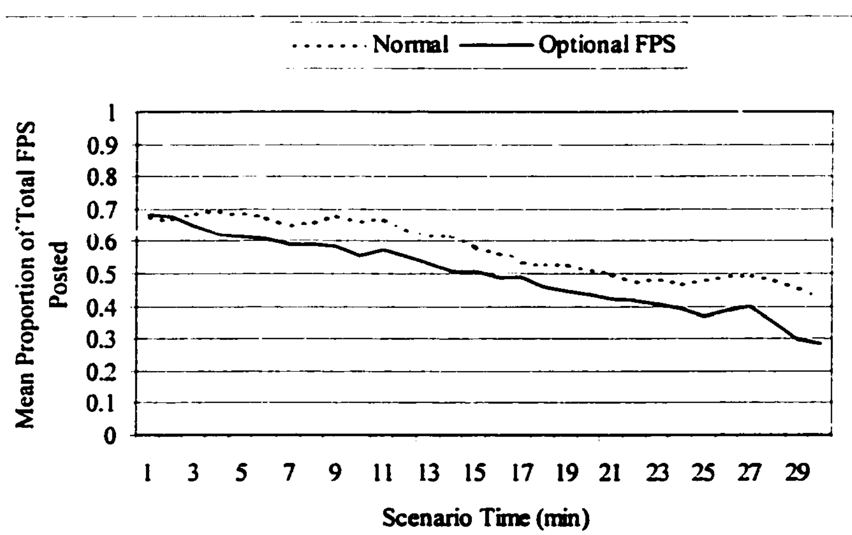


Figure 2. Mean proportion of total FPS posted by condition and scenario time for Individuals in High Altitude sectors.

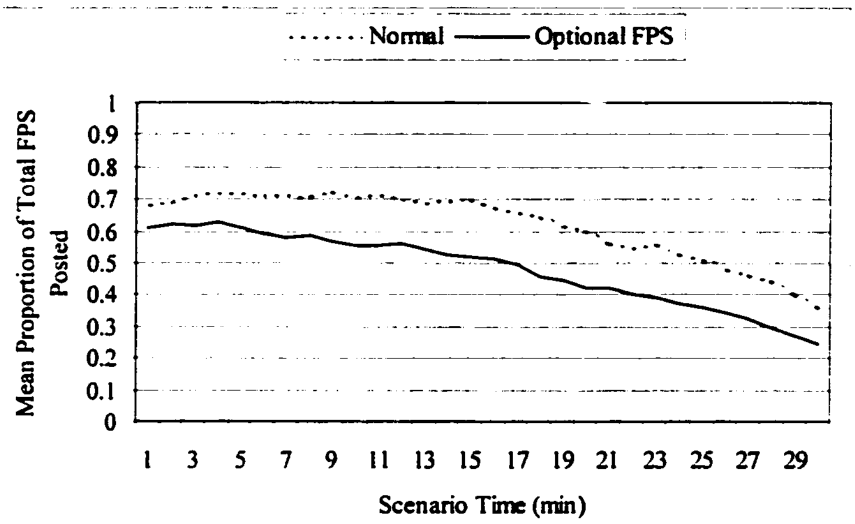


Figure 3. Mean proportion of total FPS posted by condition and scenario time for Teams in Low Altitude sectors.

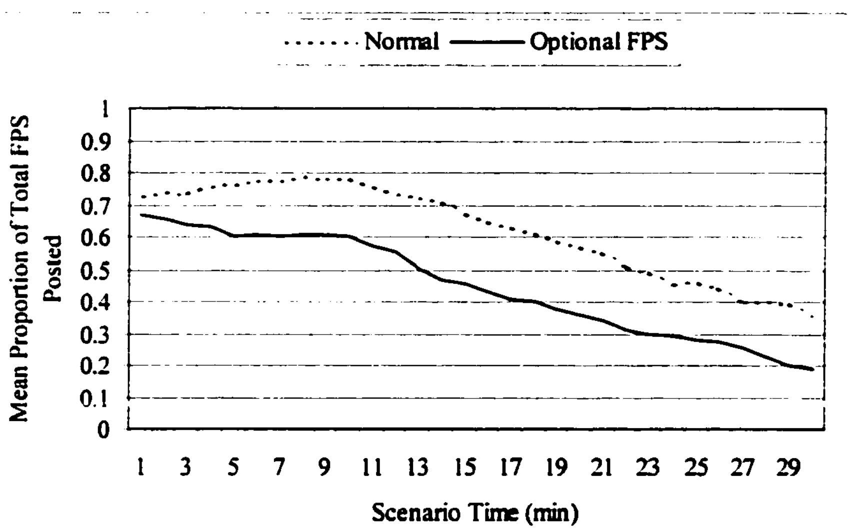


Figure 4. Mean proportion of total FPS posted by condition and scenario time for Teams in High Altitude sectors.

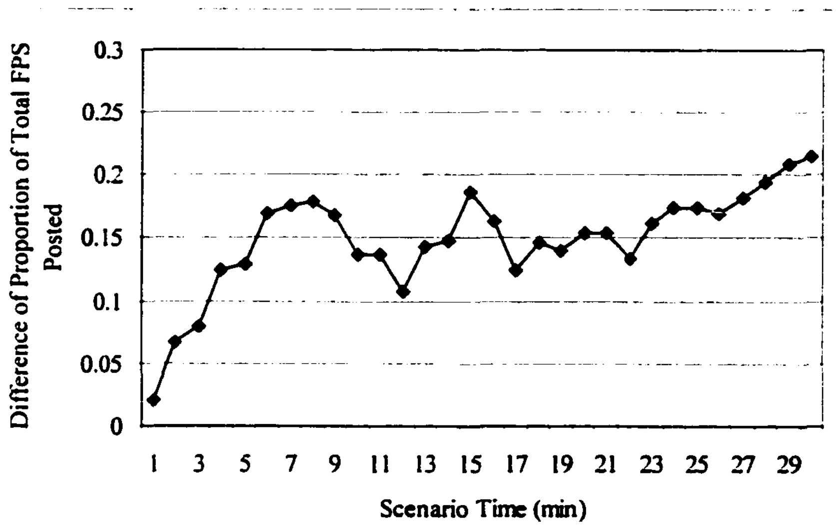


Figure 5. Difference between proportion of total FPS posted in Normal and Optional FPS conditions for Individuals in Low Altitude sectors.

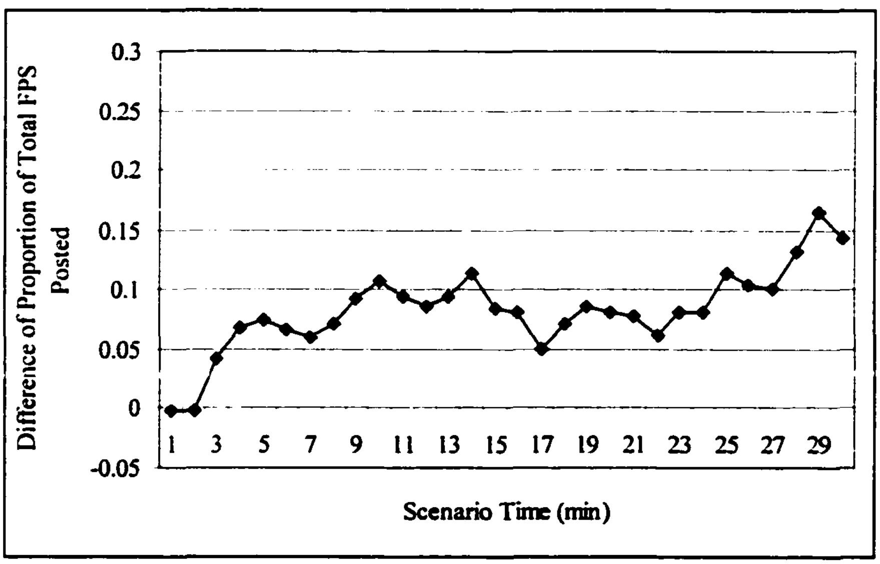


Figure 6. Difference between proportion of total FPS posted in Normal and Optional FPS conditions for Individuals in High Altitude sectors.

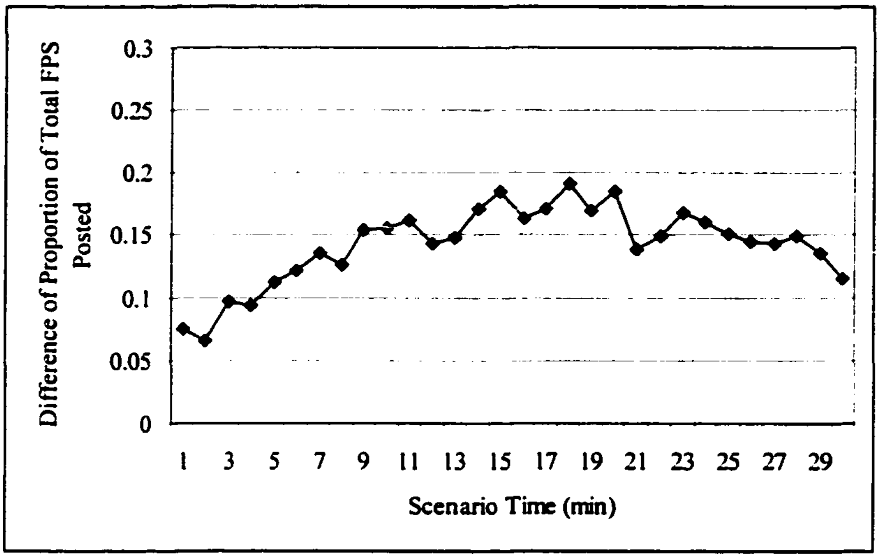


Figure 7. Difference between proportion of total FPS posted in Normal and Optional FPS conditions for Teams in Low Altitude sectors.

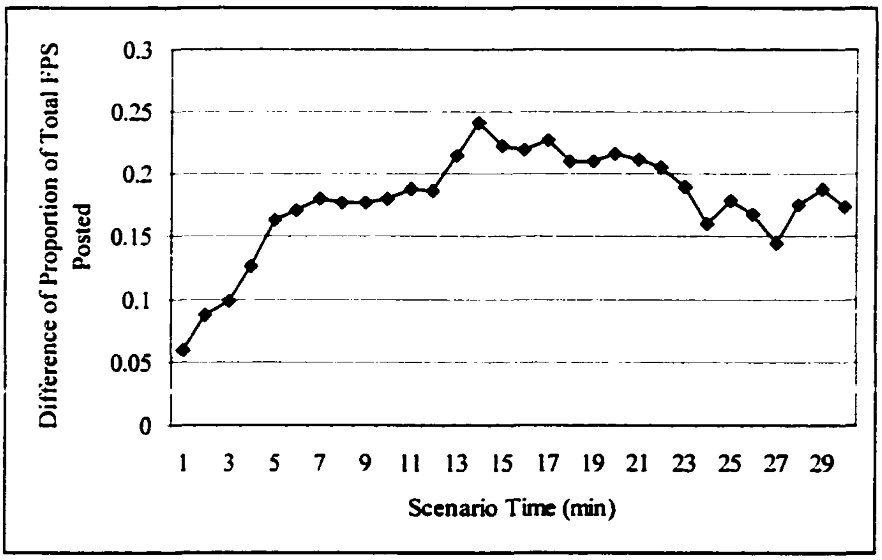


Figure 8. Difference between proportion of total FPS posted in Normal and Optional FPS conditions for Teams in High Altitude sectors.

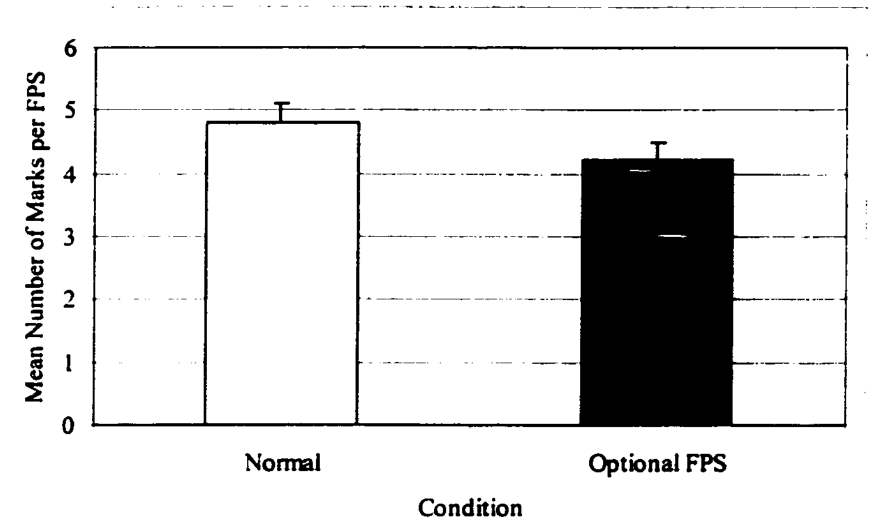


Figure 9. Mean number of marks per FPS by condition for Individuals in Low Altitude sectors.

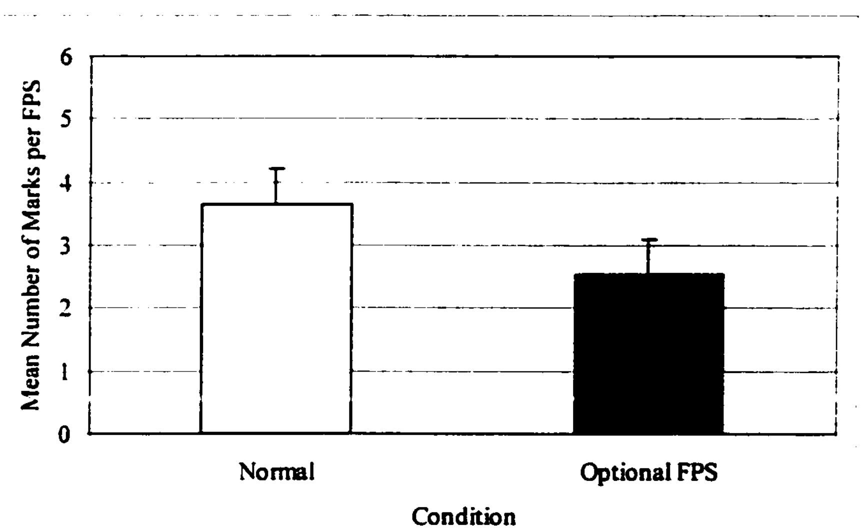


Figure 10. Mean number of marks per FPS by condition for Individuals in High Altitude sectors.

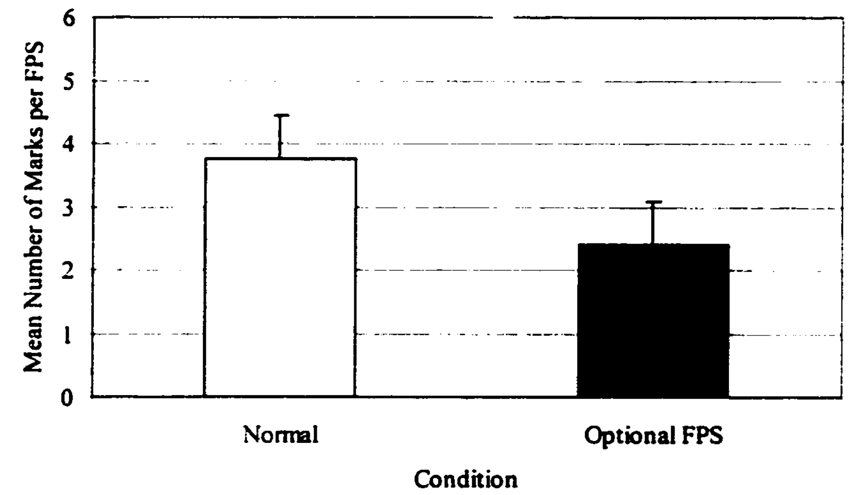


Figure 11. Mean number of marks per FPS by condition for controller Teams in Low Altitude sectors.

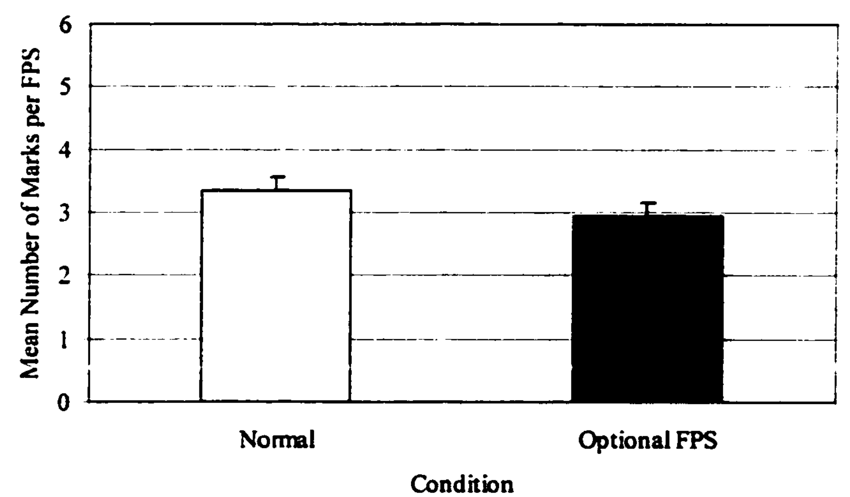


Figure 12. Mean number of marks per FPS by condition for controller Teams in High Altitude sectors.

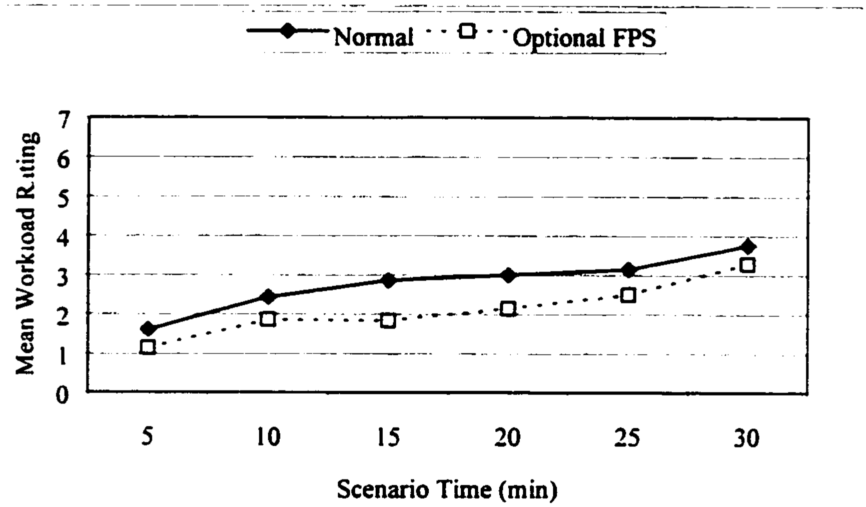


Figure 13. Mean workload (WAK) rating by condition and scenario time for Individual controllers in Low Altitude sectors.

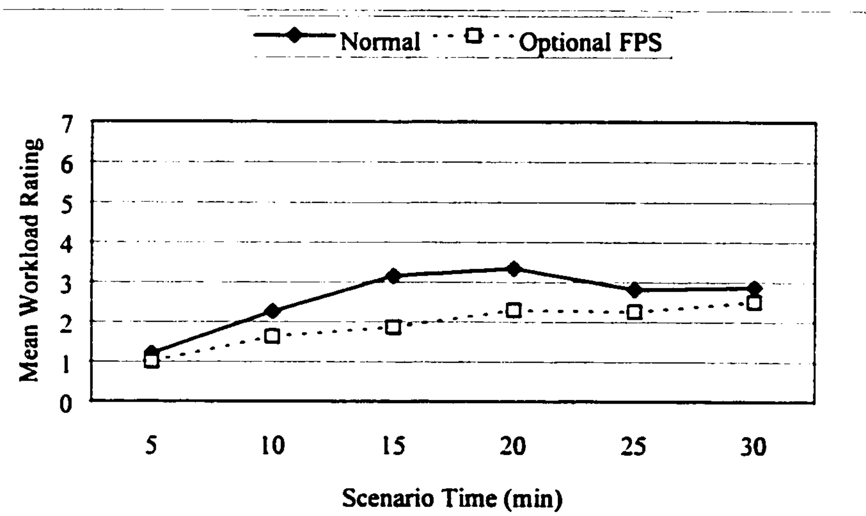


Figure 14. Mean workload (WAK) rating by condition and scenario time for Individual controllers in High Altitude sectors.

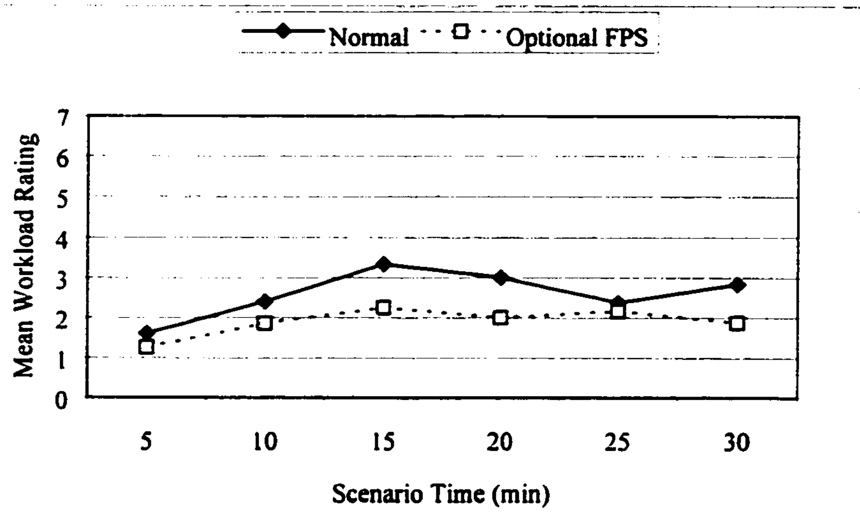


Figure 15. Mean workload (WAK) rating by condition and scenario time for Team R-side controllers in Low Altitude sectors.

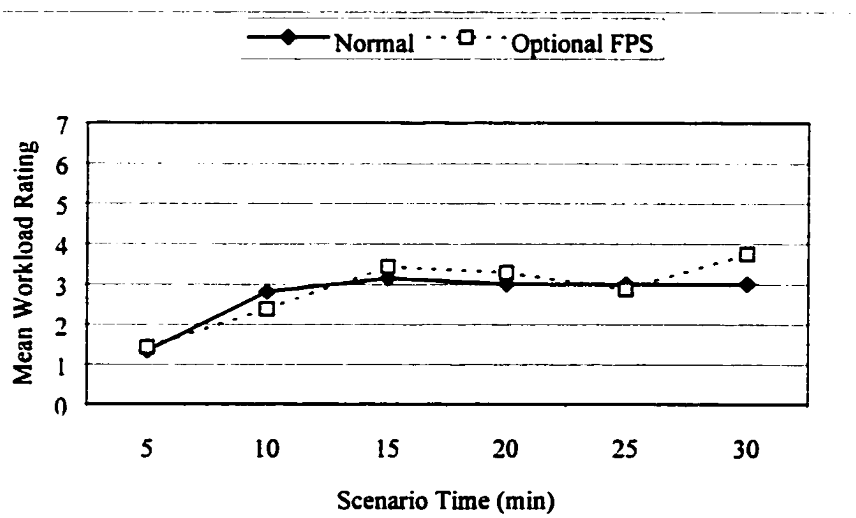


Figure 16. Mean workload (WAK) rating by condition and scenario time for Team R-side controllers in High Altitude sectors.

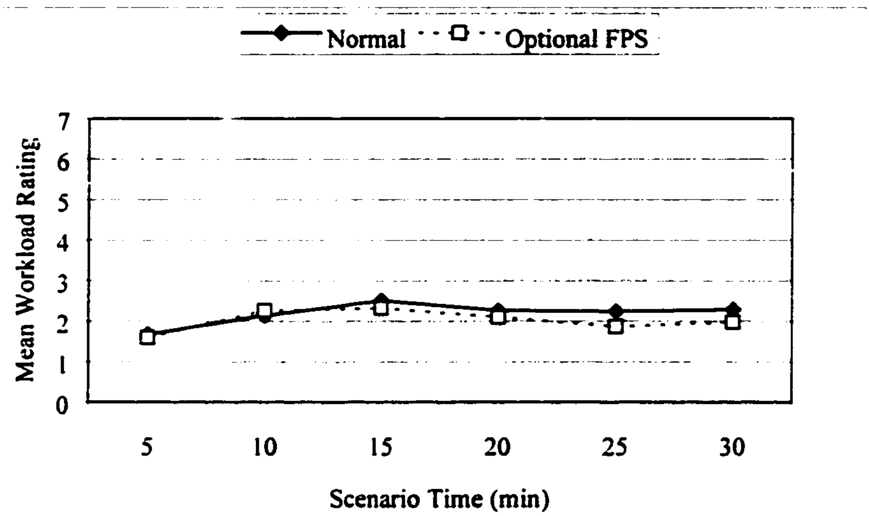


Figure 17. Mean workload (WAK) rating by condition and scenario time for Team D-side controllers in Low Altitude sectors.

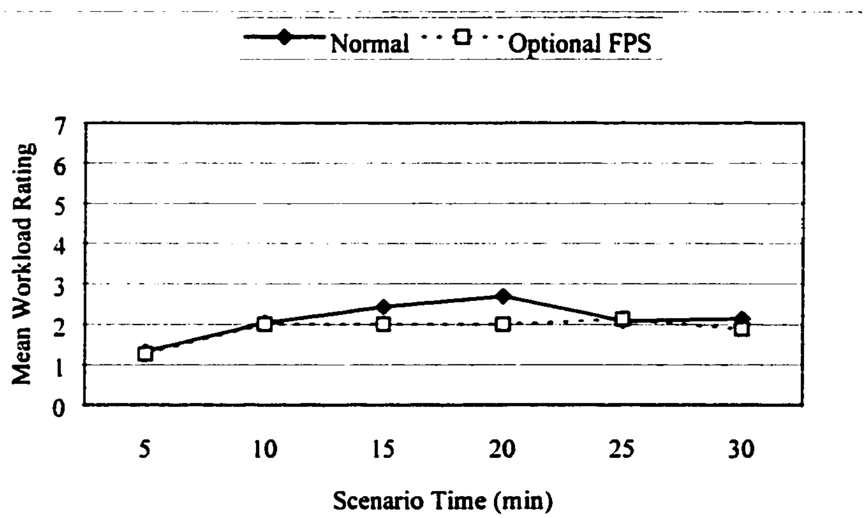


Figure 18. Mean workload (WAK) rating by condition and scenario time for Team D-side controllers in High Altitude sectors.

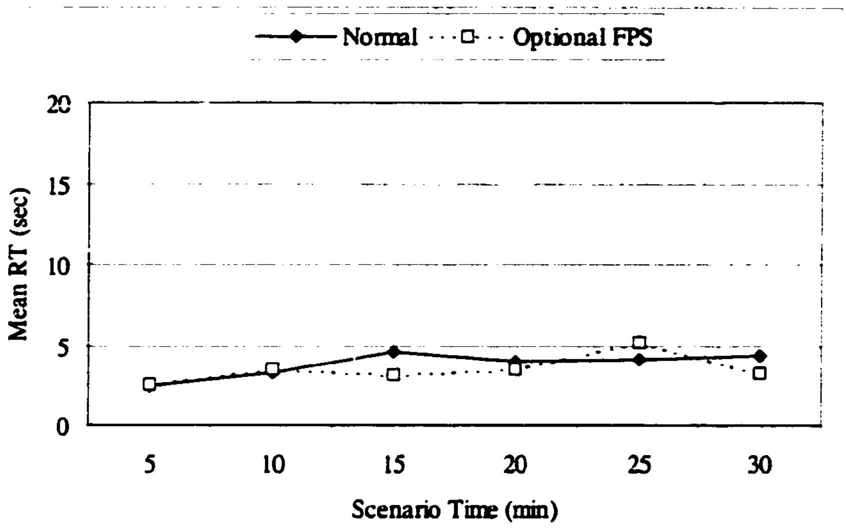


Figure 19. Mean time to respond to workload (WAK) rating by condition and scenario time for Individual controllers in Low Altitude sectors.

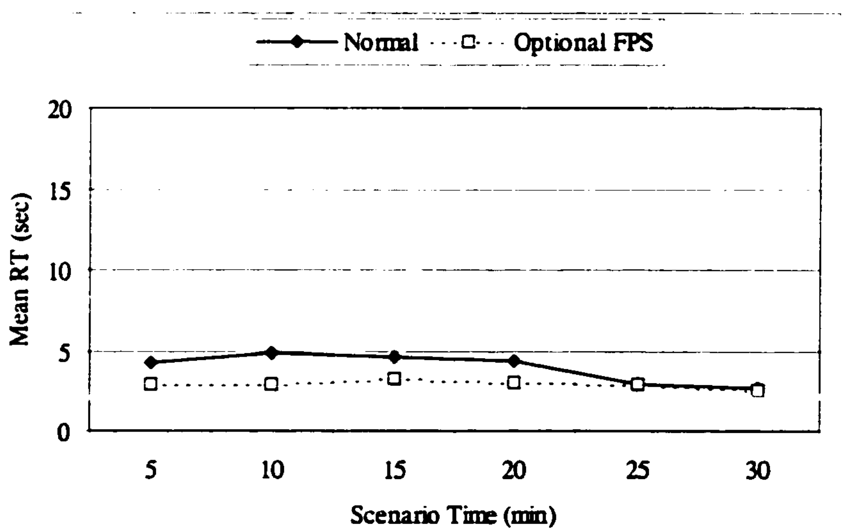


Figure 20. Mean time to respond to workload (WAK) rating by condition and scenario time for Individual controllers in High Altitude sectors.

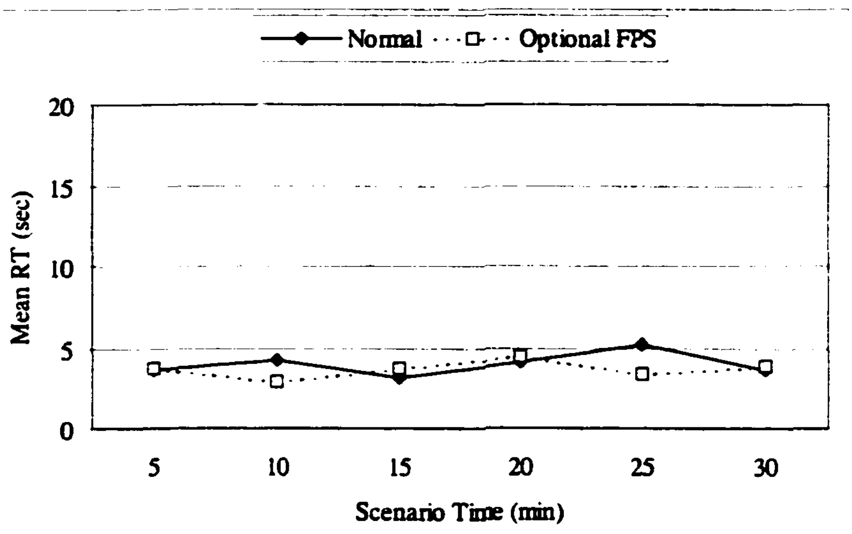


Figure 21. Mean time to respond to workload (WAK) rating by condition and scenario time for R-side controllers in Low Altitude sectors.

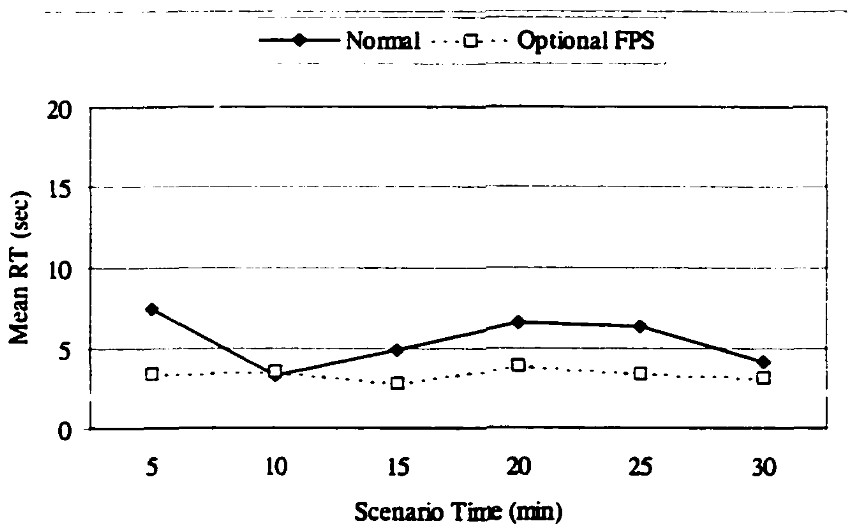
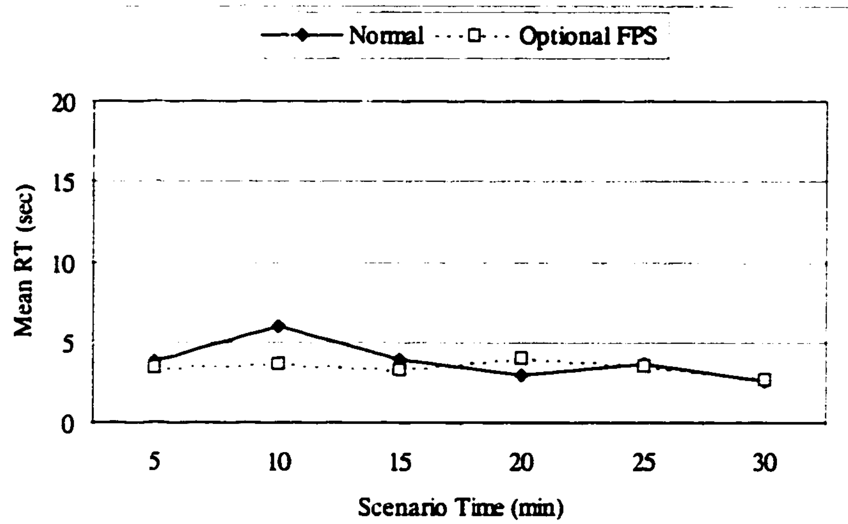
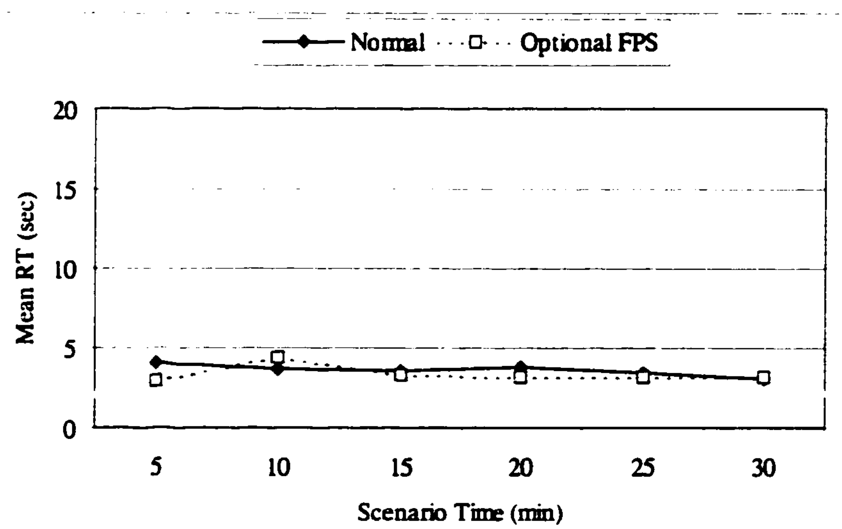


Figure 22. Mean time to respond to workload (WAK) rating by condition and scenario time for R-side controllers in High Altitude sectors.



<u>Figure 23.</u> Mean time to respond to workload (WAK) rating by condition and scenario time for D-side controllers in Low Altitude sectors.



<u>Figure 24.</u> Mean time to respond to workload (WAK) rating by condition and scenario time for D-side controllers in High Altitude sectors.

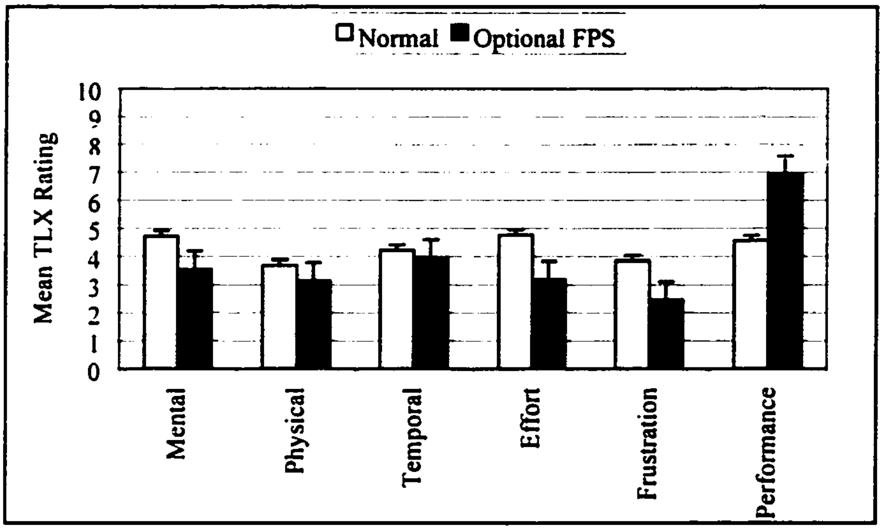


Figure 25. Mean TLX ratings by condition for Individual controllers in Low Altitude sectors.

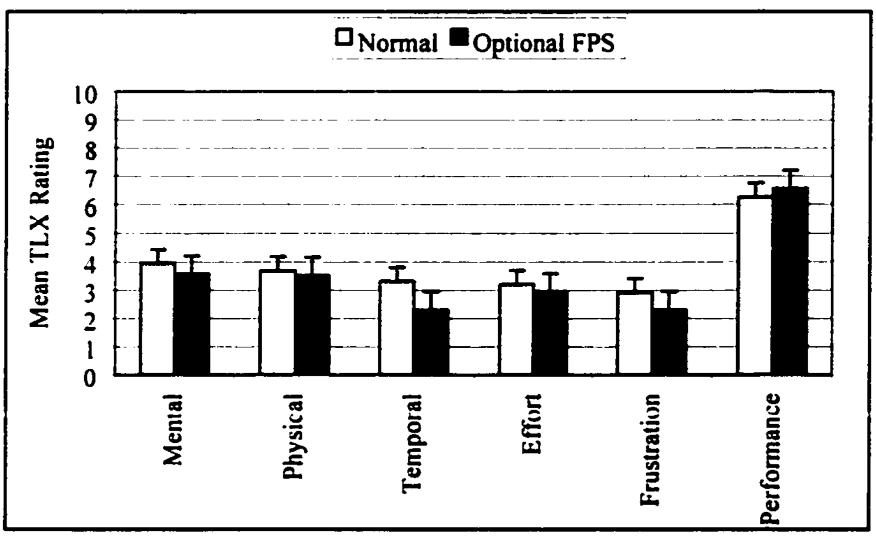


Figure 26. Mean TLX ratings by condition for Individual controllers in High Altitude sectors.

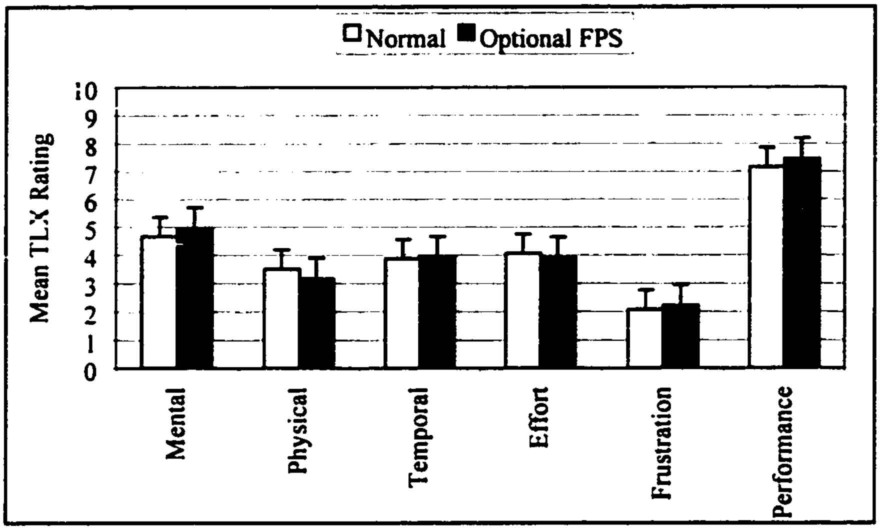


Figure 27. Mean TLX ratings by condition for Team R-side controllers in Low Altitude sectors.

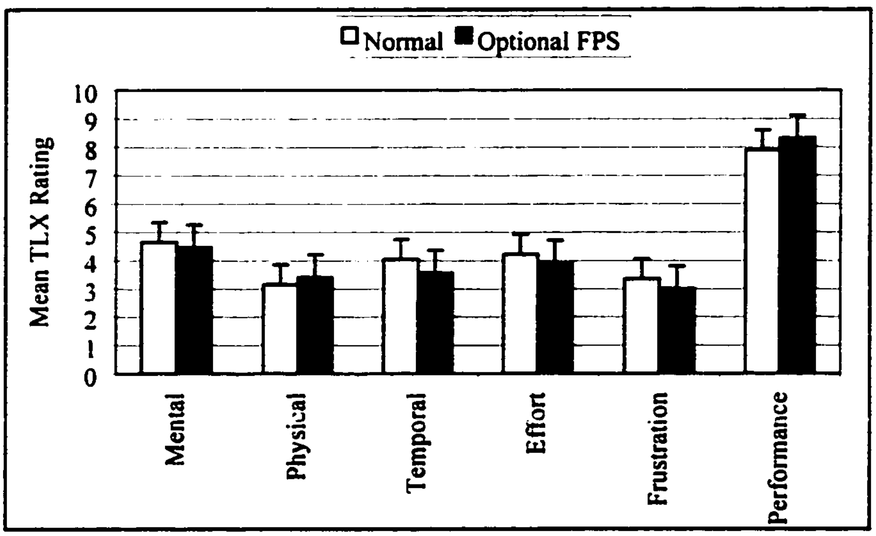


Figure 28. Mean TLX ratings by condition for Team R-side controllers in High Altitude sectors.

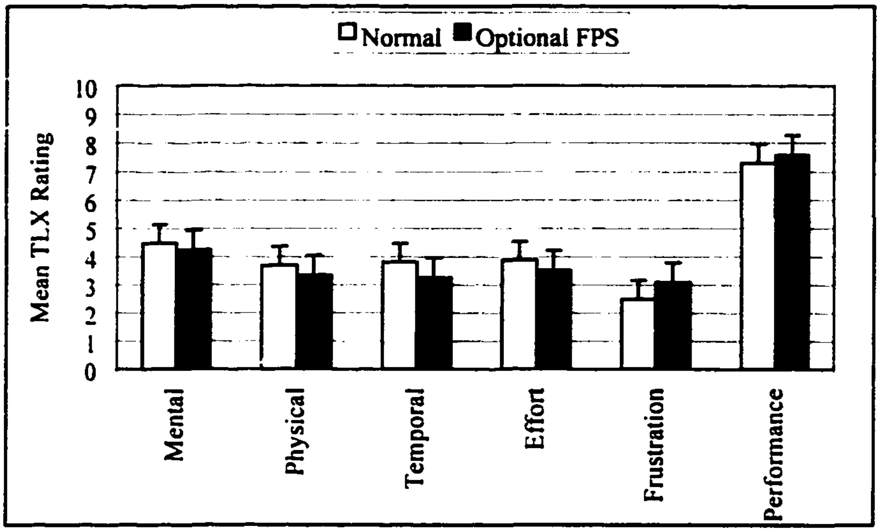


Figure 29. Mean TLX ratings by condition for Team D-side controllers in Low Altitude sectors.

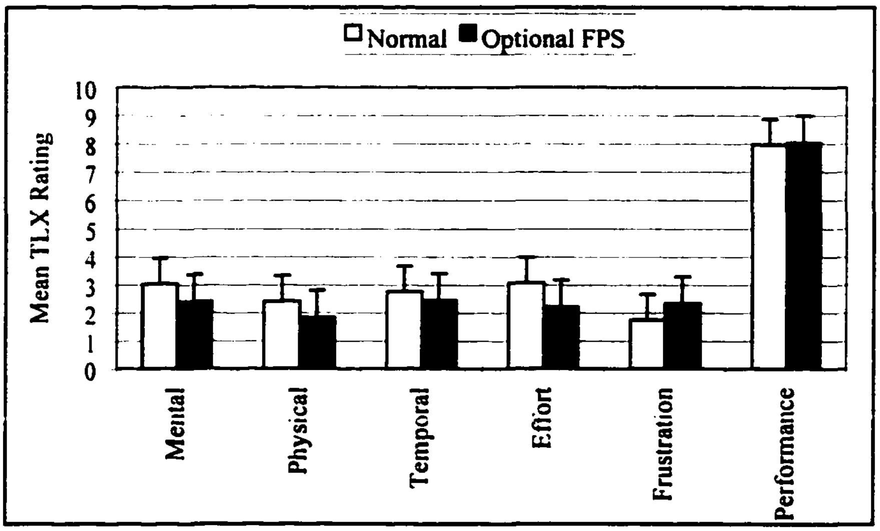


Figure 30. Mean TLX ratings by condition for Team D-side controllers in High Altitude sectors.

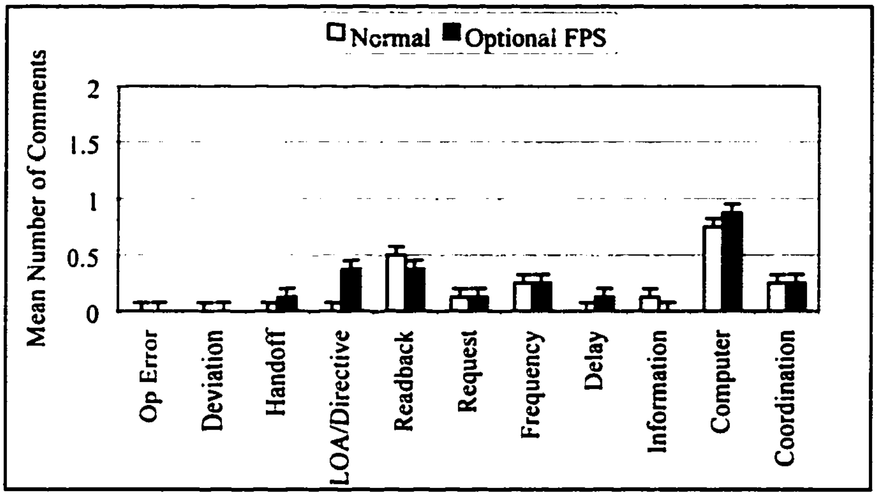


Figure 31. Mean number of event occurrences recorded by SME regarding Individual controller performance in Low Altitude sectors.

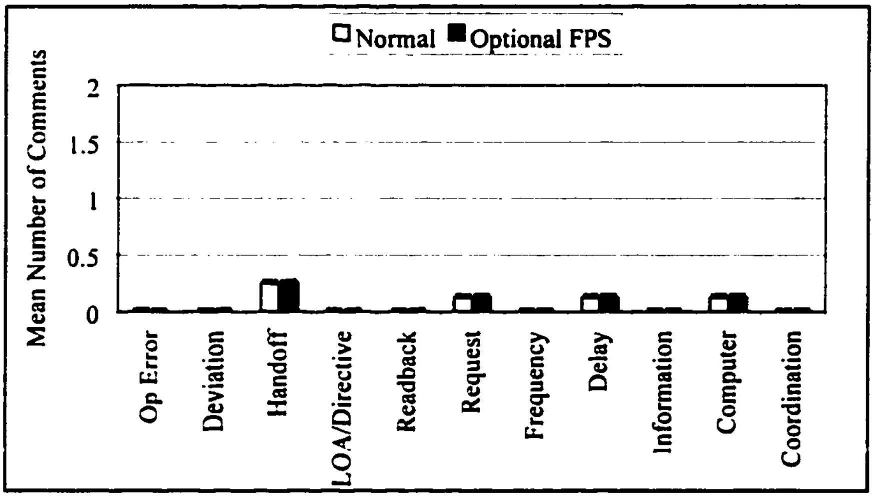


Figure 32. Mean number of event occurrences recorded by SME regarding Individual controller performance in High Altitude sectors.

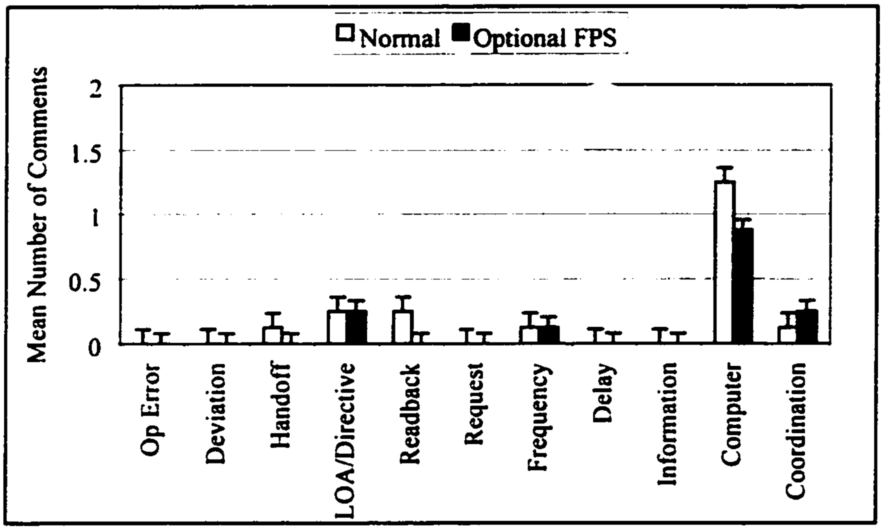


Figure 33. Mean number of event occurrences recorded by SME regarding controller Team performance in Low Altitude sectors.

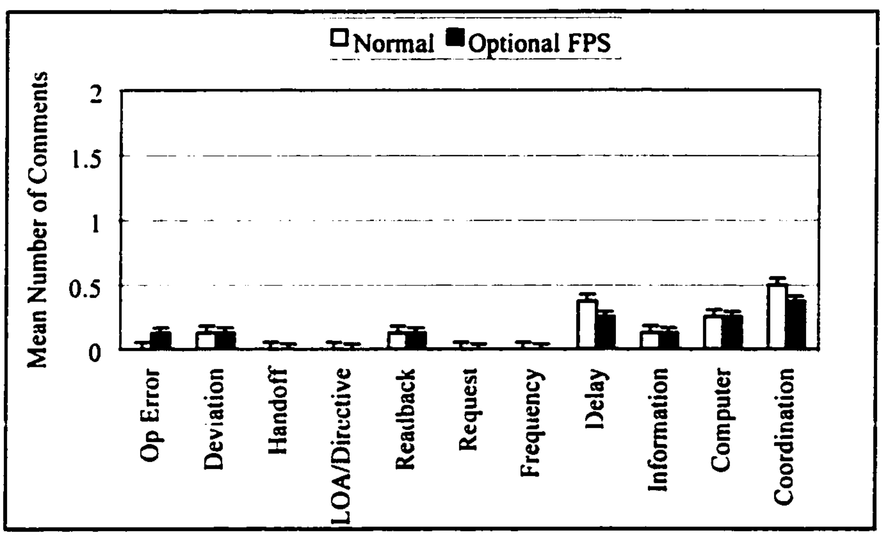


Figure 34. Mean number of event occurrences recorded by SME regarding controller Team performance in High Altitude sectors.

Appendix A

Proposed change to Order 7210.3M Part 2. Chapter 8. Section 1. Paragraph 8-1-6. Flight progress strip usage/marking procedures.

a. Flight progress strips will continue to be posted, marked, and updated in accordance with the National directives until radar contact and communications are established and accepted.

b. Once this has been achieved, the radar controller (or, if so designated, the manual controller) may remove the strip from the board. If the radar controller (or, if so designated, the manual controller) elects to keep the strip at the sector, a check mark may be placed in box 21-24 to indicate that further strip marking is unnecessary.

c. The sector team is responsible for all information contained on the flight progress strip. Standard strip marking is optional for both controllers. However, if the radar controller chooses to utilize standard strip marking, the associate controller will support and comply with that request. If the radar controller does not choose to utilize standard strip marking, nothing in this procedure precludes the associate controller from utilizing standard strip marking.

d. Partial recording of control information deemed useful to the sector operation is permitted.

e. Strips on aircraft "pointed out" to the sector may be check marked even though communications are never established.

f. The following flight progress strips are to remain posted and standard strip marking used:
   (1) Any aircraft you cannot reasonably expect to remain in radar contact.
   (2) Aircraft transitioning from radar to non-radar environment.
   (3) Aircraft requiring special handling, i.e., emergencies, radio failures, etc. (Note: Standard strip marking will begin when the need for special handling is identified.)
   (4) All non-radar flights.
   (5) All flights transitioning from automated to non-automated modes of operation.
   (6) All flights that are issued holding instructions.

g. Departures and proposals are to be considered non-radar until radar contact has been established and accepted in accordance with FAAH 7110.65.

h. Any control action not recorded via landline, frequency, or computer entry must be marked on the appropriate strip to record the action.

i. In the event of a back-up system failure, the strips will be posted and marked in accordance with National and local directives. These systems would include any back-up radar or malfunctioning recording system and when operating in DARC only.

j. The controller will continue to be responsible for the control of the aircraft and coordination of information as prescribed in FAAH 7110.65.

k. Blank note pads will be available at every sector for the controller's use.
l. Standard strip marking must be accomplished when training is in progress. This may be discontinued with the consent of the training team.

Appendix B

Biography

1) At how many En Route Centers have you worked?
_________ (number of centers)

Please list the centers you have worked at beginning with the most recent:

1) ______________________________
2) ______________________________
3) ______________________________
4) ______________________________

2) How long have you worked in your current area?
______ years _______ months

3) How long have you worked at your current ARTCC?
______ years _______ months

4) How many years and months (total) have you worked at an En Route Center as an FPL? ______ years _______ months

5) Please indicate your total number of years as a controller.
______ years _______ months

6) Please indicate all operations in which you have been an FPL (check all that apply)

______ En Route
______ Terminal
______ Flight Service Station
______ Other (please list)

7) Please check below any area in which you served for six months or more. Please indicate whether you were a manager (M) or a specialist (S). (check all that apply)

______ Flow Control
______ Traffic Management Unit
______ Supervisor at En Route
______ Training Department
______ Automation Specialist
______ Quality Assurance Specialist
______ Area Officer
______ Regional Officer
______ National Headquarters

8) When were you last certified or re-certified?

______ month _______ year

OPTIONAL: Please provide the following:

1) Age _____

Gender _____

Appendix C

General descriptions for airspace sectors used.

Cleveland ARTCC (ZOB)

Lichtfield (LFD) Sector – low altitude (ZOB 7220.2B, Appendix 1)

The LFD Sector is a major feeder sector for arrival traffic to the DTW area. Most of this traffic comes in from the west, southwest, and the south. LFD Sector also works TOL arrivals and departures to or from the west and northwest.

Hudson (HUD) Sector – high altitude (ZOB 7220.2A, Appendix 2)

The HUD Sector is an intermediate high altitude sector that has two (2) main flows of traffic: One is primarily eastbound from the CRL VOR. The other flow is westbound via J36.FNT. There is additional eastbound traffic on J70. With the HUD Sector is an air refuel route, AR206H, which is located along J36.

Jacksonville ARTCC (ZJX)

Florence (FLO) Sector – low altitude (ZJX 7220.4H, Appendix 3)

This sector shall include all airspace from the surface up to but not including FL240.

Brewton Sector – high altitude (ZJX 7220.4H, Appendix 5)

This sector shall include all airspace from the surface up to but not including FL240.

Appendix D

Workload Assessment Keypad (WAK) Instructions

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. Every five minutes the WAK device will emit a brief tone and the buttons will illuminate. The buttons will remain lit only for a limited amount of time. Please tell us how hard you are working by pushing a button on the WAK numbered from 1 to 7.

I will review what these buttons mean in terms of your workload. At the low end of the scale (1, 2), your workload is low - you can accomplish everything easily. As the numbers increase, your workload is getting higher. Numbers 3 and 4 represent increasing levels of moderate workload where effort is low and the chance of errors are low but steadily increasing. Numbers 5 and 6 reflect relatively high workload where effort is high and there is some chance of making errors. At the high end of the scale is the number 7, which represents a very high workload, where it is likely that you will have to leave some tasks unfinished.

All controllers, no matter how proficient or experienced, will be exposed at one time or another to all levels of workload. It does not detract from a controller's professionalism when he indicates that he is working very hard or that he is hardly working. Feel free to use the entire scale and tell us honestly how hard you are working. You will have 20 seconds to respond. However, do not sacrifice the safe and expeditious flow of traffic in order to respond to the workload rating form. Remember, your rating should reflect how much workload you are experiencing during the instant when you are prompted to make the rating.

Do you have any questions about using the WAK device?

Appendix E

TLX Instructions

We are interested in finding out your perception of how difficult the task is and how well you perform the task. Our objective is to measure your perceived "workload" level. The concept of workload is composed of several different factors. Therefore, we would like you to tell us about several individual factors rather than one overall workload score.

Here is an example of the rating scales. As you can see, there are six scales on which you will be asked to provide a rating score: *mental demand, physical demand, temporal demand, effort, frustration, and performance.*

<u>Rating Scales</u>

Mental demand refers to the level of mental activity like thinking, deciding, and looking that was required by the task. You will rate this scale from low to high.

Physical demand involves the amount of physical activity required of you, such as controlling or activating.

Temporal demand refers to the time pressure you experienced during the task. In other words, was the pace slow and leisurely or rapid and frantic? If the pace was rapid and frantic you are experiencing high temporal demand.

Effort refers to how hard you worked (both mentally and physically) in order to achieve your level of performance.

Frustration level refers to how secure and relaxed versus stressed and discouraged you felt during the task. If you feel secure and relaxed, you have low frustration.

Performance level refers to you perception of your own performance level. Your rating here should reflect your satisfaction with your performance in accomplishing the goals of the task.

<u>Making your response</u>

You should indicate your rating by placing an 'X' on the line adjacent to the item.

For example, if you want to give a high rating of stress factor, place an 'X' to the right of the half-way mark. The higher the stress rating, the closer the 'X' should be to "HIGH". In contrast, if your stress rating is low, you would place an 'X' closer toward the "LOW" end of the line. Likewise, if the stress rating is average, place an 'X' in the center of the line.

Please give your responses thoughtful consideration, but do not spend too much time deliberating over them. Your first response will probably accurately reflect your feelings and experiences.

Appendix F – SME rating form

Behavioral and Event Checklist

Event	**Aircraft identity**	**Totals**
Operational Errors (Describe briefly in this column)	(Write both call signs in one box)	
1.		
2.		
3.		
Operational Deviations/SUA violations (Describe briefly in this column)	(Write call sign in each box)	
1.		
2.		
3.		
4.		
Behavior	**Number of events**	**Totals**
Failed to accept, initiate handoff		
LOA/Directive Violations		
Readback/Hearback errors		
Failed to accommodate pilot request		
Made late frequency change		
Unnecessary delays		
Asked pilot for information available from computer or strip		
Incorrect information in computer		
Failed to pass headings, speeds or coordinate pointouts		

Appendix G - Team Communication Effectiveness Rating Form

Effectiveness of Team Communications

Sector type: High Low	SME ID:
Condition: Bascline Experimental	Scenario ID:
Facility: ZOB	Participant ID:

Activity	Effective	Ineffective	N/A
Each controller was aware of actions other had taken			
Communicated verbally with team member			
Pointed at PVD to communicate with team member			
Pointed at strips to communicate with team member			
Duplicated efforts—repeated actions already taken by other team member			
	True	Not true	N/A
Lack of communication affected safety of operations			
Lack of communication affected efficiency of operations			

Notes (comment on specific events):

Appendix H - Position Relief Briefing Rating Form

Position Relief Briefing Checklist

To be completed by SME or ATC Task Force Member

Individual / Team Sector type: High Low	SME ID:
Condition: Baseline Experimental	Scenario ID:
Facility:	Participant ID:

Please make a mark under the appropriate column if the participant did not cover the items along the left side of the form during the position relief briefing.

Items briefed	Individual	Team R-side	Team D-side
Navaids			
Equipment			
Radar			
Airports			
Weather			
Flow Control			
Special Use Airspace			
Traffic			

Please choose one of the following and comment below, if desired.

In your opinion, was the position relief briefing negatively affected by the absence of strips of the lack of strip marking?

a. No

b. Yes, it was affected by the absence of strips

c. Yes, it was affected by the lack of strip marking

d. Yes, it was affected by both b and c.

Appendix I - Procedures Summary Sheet

FPS Procedures

Post and mark FPS until radar contact and communications are established and accepted.

Must record on FPS any action not recorded elsewhere.

To indicate no further marking required, check box 21-24.

If FPS no longer needed, may remove from board.

Must post and mark FPS for aircraft if:

1) Radar contact may be lost
2) Transitioning from radar to non-radar
3) Special handling required
4) Non-radar flight
5) Transitioning from auto to non-auto mode
6) Holding instructions are issued

Appendix J

Applications of Item Response Theory

Beyond the standard analyses that use classical test procedures such as ANOVA, t-tests, and chi-square, an examination of the subjective measures of workload as measured by the WAK and the TLX was conducted using item response theory (IRT). If IRT can be applied to these data, then a more thorough and valid interpretation of the examined data may be obtained. Because IRT is not well known and has not been applied to data such as presented here, also investigated was the value of IRT analyses with respect to data such as that obtained from the WAK and TLX.

Like most data in both basic and applied psychological research, analysis of subjective measures of workload typically adheres to the time-tested utility of classical test theory. In contrast, IRT is not widely used. Perhaps IRT is not often used because it is not well known, it is relatively complicated, or the computer tools for conducting these types of analyses have not been well formed. Perhaps IRT is believed only to be of benefit to a particular type of psychology such as intelligence testing or selection for industrial/organizational settings. I believe none of these reasons are sufficient for neglecting the use of IRT.

Item response theory is a type of latent variable analysis that is able to overcome some of the major pitfalls of classical test theory. In classical test theory, an estimate of the theoretical true score (T) is obtained by measuring some observed score (X) and a component of measurement error (E) such that the model T = X + E results. However, this model is tautological and cannot be disproved. Furthermore, in assessing a true score, classical test theory must assume a response by occasion interaction. In other words, an observed response is always a matter of the underlying construct that is being measured in conjunction with a multitude of other factors including item and temporal characteristics of the measure. Examinee characteristics and test characteristics cannot be separated. Classical test theory does not account for this interaction and it follows that estimated true scores can only be interpreted in the context in which those scores were obtained. The practical consequence is that item characteristics are group-dependent and it becomes difficult to compare test items that are obtained from widely different groups. For example, items of a proficiency exam would be group-dependent if they were field-tested in the Fall and then the actual test was administered in the Spring. Similarly, classical test theory results in examinee scores that are test-dependent and it therefore becomes difficult to compare examinees who take different tests. Each score from a different test (or the same test) contains different amounts of error even if the various tests are parallel. The fact that two scores are not equally reliable is demonstrated by the lack of precision for a score of zero. Although a score of zero suggests low ability, how low that ability actually is cannot be determined.

A final caveat of classical test theory is that estimates of true scores are many times an amalgamation of a number of observed scores, all of which contribute equally to a total observed score. Classical test theory does not address the varying difficulty, or discriminability, of items when estimating an overall true score. Classical test theory is, therefore, test-oriented rather than item-oriented in that it focuses on the overall test score rather than responses to particular items. The practical consequence is that classical test theory is unable to predict how an individual or group will perform on any particular item.

Related to signal detection theory, IRT provides a model that estimates threshold, slope, and ability parameters. In contrast to classical test theory, IRT is based on models that can be tested and falsified. Any resultant model can be supported or rejected by testing for goodness-of-fit.

As mentioned previously, the IRT model assesses a latent variable, namely the ability parameter. The ability parameter is theoretical and can never be known, although it can be estimated by using the threshold and slope parameters. The ability parameter is not confined to "ability" per se, it is simply the latent variable of interest. The ability parameter could, for example, estimate need for cognition, or subjective workload. The threshold parameter indicates where on the ability dimension one must lie in order to produce a particular response to an item. The slope parameter indicates the discriminability of an item (or rater), with higher slopes indicating greater discriminability.

In contrast to classical test theory, which uses an overall score to assess ability, IRT uses both the threshold and slope parameters in conjunction with the pattern of responses to the items in order to provide an estimate of the latent ability variable. This means that even though two participants may have the same overall test score, their differing patterns of responses to items of varying difficulty and discriminability will yield different estimates of the latent variable being measured (the ability parameter) when IRT is used. The estimate of the latent variable, or ability parameter, is no longer test-dependent because scores are estimated at the level of the item rather than the level of the test. A measure of precision is also provided for each ability estimate.

A final advantage of IRT is that it can test whether a rating-response model, or a graded-response model (Samejima, 1969) best describes the data. These two models are important to distinguish when using likert-type rating scales. The rating-response model states that the psychological distance between the available ratings, say 1 to 10, remains the same across various items or repeated measures of the same item. The graded-response model, on the other hand, does not assume a fixed psychological rating system given different items. The more complicated graded-response model allows for the fact that a rating scale may not remain stable across different items or multiple ratings and estimates additional threshold parameters to describe each item independently. If the graded-response model fits the data better than a rating-response model, then the estimates of the ability parameter will gain precision and validity by accounting for item characteristics.

Modeling scores at the item-level by using IRT rather than at the test-level by using classical test theory results in a number of benefits. Benefits include a measure of precision for each ability score, estimates of reliability without the assumption of parallel tests, item characteristics that are not group-dependent, and ability estimates, or scores, that are not group-dependent.

To achieve the benefits, IRT rests on three basic assumptions. First, the assumption of unidimensionality states that only one dominant ability is measured by the test. The assumption of unidimensionality is probably never strictly met due to cognitive, personality, and test taking factors. Additionally, IRT does not assume that ability is inherent or unchanging, but rather that scores will change over time.

Local independence is the second assumption of IRT. The assumption of local independence states that responses to items are independent when one controls for ability. In other words, responding to a particular item does not affect how one will respond to other items. The answer to one item does not provide any clues to the answers to any other items. The local independence assumption is also related to the property of invariance. That is, item characteristics are invariant and are not affected by the ability of any particular group of examinees.

The third assumption states that an item characteristic curve (ICC) is a true representation of the relationship between unobservable traits and observable item responses. This assumption results from the consideration of the unidimensionality and local independence assumptions. The ICC may be described by a one-, two-, or three-parameter logistic model as a monotonically increasing function. Only the one- and two-parameter models will be discussed because the three-parameter model only deals with situations in which guessing may be a factor.

The one-parameter model is described by the following equation:

$$P_i(\theta) = e^{(\theta - b_i)} / [1 + e^{(\theta - b_i)}] \quad i = 1, 2, \ldots, n$$

where $P_i(\theta)$ is the probability that a random examinee with ability θ answers item i correctly; b_i is the difficulty parameter for item i; n is the number of items in the test; and e is a constant, transcendental number equal to 2.718. The difficulty parameter, b_i, is the point on the ability scale where the probability of a correct response is 0.5. The estimates of ability, θ, are transformed to have a mean of 0 and a standard deviation of 1. The transformation results in values of b_i ranging from –2.0 to +2.0. The one-parameter model suggests that various items are equally capable of discriminating between different abilities but each item varies in difficulty. In other words, all items are equally discriminating and the item difficulty parameter is the only factor influencing performance in the one-parameter model.

The two-parameter logistic model based on work by Lord (1952) and Birnbaum (1968) builds on the one-parameter model by accounting for each item's power to discriminate between levels of ability in addition to each item's difficulty. The two-parameter model results from the following equation:

$$P_i(\theta) = [\ 1 + e^{-Da_i(\theta - b_i)}]^{-1} \quad i = 1, 2, \ldots, n$$

where a_i is the item discrimination parameter and is proportional to the slope of the ICC at b_i; and D is a constant scaling factor equal to 1.7. Theoretically, a_i may range from zero to $+\infty$, but generally ranges from 0 to 2.

Application of IRT to WAK Data

The WAK data were analyzed using a two-parameter graded-response model. Due to the small number of participants in each condition, participant characteristics were not modeled. However, the item slopes and thresholds were modeled for each experimental condition, which resulted in ability estimates, that is, workload estimates, for each participant. Values of the –2 Log Likelihood deviance statistic ranged from 73.82 to 102.70 suggesting some lack of model fit. These values may decrease as more observations are used to estimate the slopes and thresholds.

Workload estimates for the Normal and Optional FPS conditions were compared to see if the estimates differed. Comparisons were made within each group of participants (Individuals/R-side/D-side and High/Low altitude sector) by averaging over each participant's estimate of workload by condition and then testing the means with a dependent-measures t-test with $\alpha = .05$. None of the statistical analyses were significant. Typically, a chi-square test is used for such a comparison but the small number of participants in this study did not provide sufficient degrees of freedom. Although these results should be interpreted with caution due to the small sample size, differences in workload estimates between the Normal and Optional FPS conditions were not detected. Mean slope and threshold estimates by condition for each group of participants are shown in Table 7.

Table 7. Mean (standard deviation) Slope and Threshold estimates by condition for the WAK measure.

	Slope		Threshold	
	Normal	Optional FPS	Normal	Optional FPS
Ind Low	0.939 (.185)	0.967 (.191)	-0.645 (.956)	0.526 (1.006)
Ind High	0.969 (.147)	1.255 (.445)	0.121 (1.016)	0.593 (.522)
R-side Low	1.137 (.270)	1.061 (.167)	0.165 (.286)	0.886 (.807)
R-side High	0.918 (.183)	1.082 (.355)	0.516 (.919)	-0.234 (.965)
D-side Low	1.149 (.328)	1.031 (.166)	0.022 (1.107)	0.490 (.349)
D-side High	0.996 (.229)	1.115 (.327)	0.154 (.417)	0.506 (.603)

In general, the slope estimates suggest that the WAK measure is discriminating between levels of workload relatively well in both the Normal and Optional FPS conditions. With exception of the R-side controllers in high-altitude sectors, threshold estimates for the Optional FPS condition are shifted to the right on the workload scale as compared to the Normal condition. This suggests that higher workload ratings are less likely in the Optional FPS condition.

Application of IRT to TLX Data

The TLX data were also analyzed using a two-parameter graded-response model. Responses to the Performance item of the TLX were reverse scored because as workload increases Performance ratings should decrease while ratings for the other five items should increase. Reverse scoring of the Performance item kept the rating scale consistent over the six items of the TLX. The item characteristics were modeled for each experimental condition and estimates of workload were obtained. The –2 log likelihood deviance statistic ranged from 92.74 to 121.24 indicating even poorer fit for the TLX data than for the WAK data. The poorer fit may be due a lack of unidimensionality in the TLX measure.

Workload estimates for the Normal and Optional FPS conditions were compared with a dependent-measures t-test ($\alpha = .05$) to see if ratings differed. Comparisons were made within each group of participants (Individuals/R-side/D-side and High/Low altitude sector) by averaging over each participant's estimate of workload by condition and then testing the means. No significant results were found suggesting that workload did not differ between the Normal and Optional FPS conditions. Table 8 shows the mean slope and threshold estimates by condition for each group of participants.

Table 8. Mean (standard deviation) Slope and Threshold estimates by condition for the TLX measure.

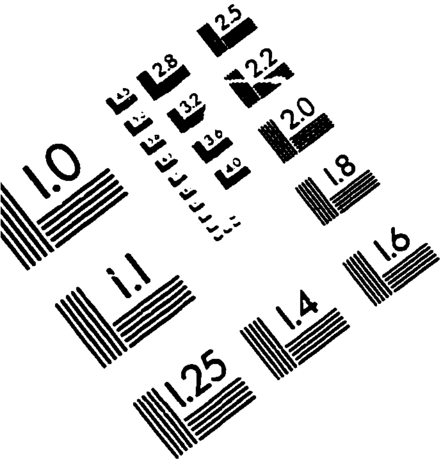
	Slope		Threshold	
	Normal	Optional FPS	Normal	Optional FPS
Ind Low	1.329 (.443)	.950 (.223)	-.315 (.421)	.442 (.322)
Ind High	1.293 (.278)	1.213 (.357)	.237 (.331)	.425 (.224)
R-side Low	1.002 (.298)	1.104 (.307)	.244 (.497)	.413 (.486)
R-side High	1.239 (.355)	1.209 (.395)	.257 (.561)	.469 (.477)
D-side Low	1.274 (.389)	1.191 (.232)	.046 (.416)	.236 (.464)
D-side High	1.282 (.409)	1.248 (.385)	.382 (.501)	.806 (.442)

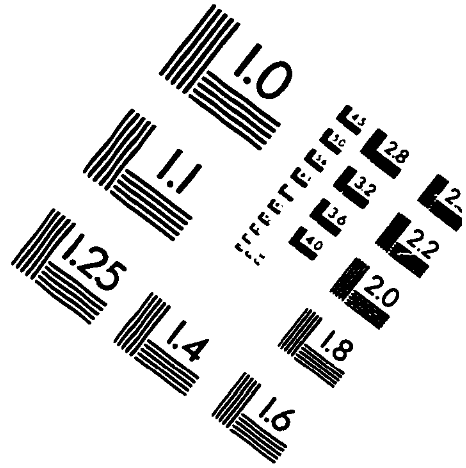
The relatively steep slope estimates indicate that the TLX measure was able to discriminate between levels of workload in both the Normal and Optional FPS conditions. Like the WAK results, all of the threshold estimates for the Optional FPS condition were higher than the threshold estimates for the Normal condition. Such a result suggests that items in the Optional FPS condition are somewhat "harder", that is, are less likely to receive high workload ratings.

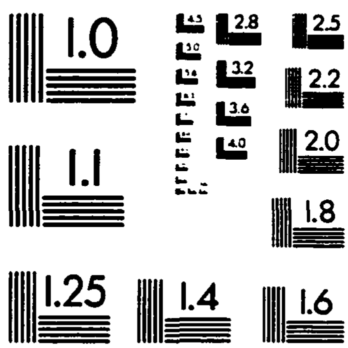
Conclusions regarding IRT

Analyses of the WAK and TLX data using IRT mirror the findings, or lack thereof, obtained from the previously conducted ANOVAs. Namely, no significant differences in workload estimates were found from the IRT analysis. The lack of a significant difference in subject workload ratings between the Normal and Optional FPS conditions may be due in part to the small number of participants in the study. Just as likely, the results suggest that there is not a large difference in workload between the two conditions. However, the threshold parameters were higher in the Optional FPS condition for both WAK and TLX data indicating that higher ratings may have been more likely to be selected in the Normal condition.

IMAGE EVALUATION TEST TARGET (QA-3)

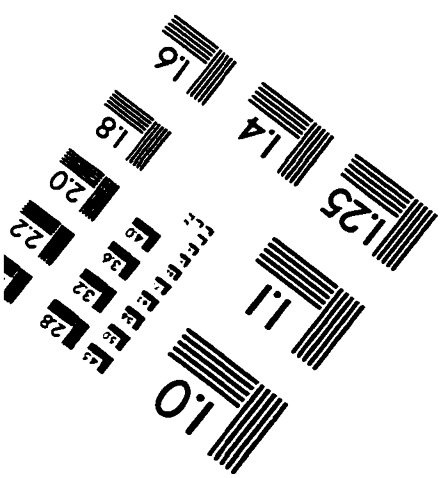


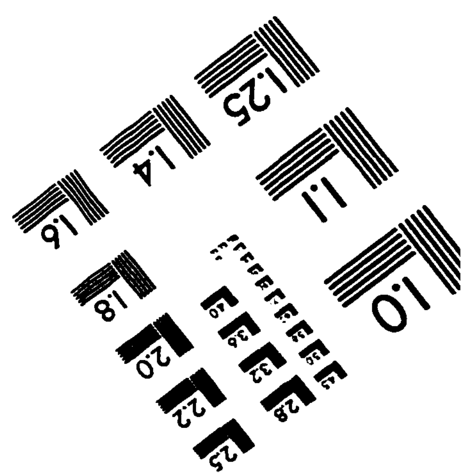




150mm

6"





APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989