DOT/FAA/AM-99/24

Office of Aviation Medicine
Washington, D.C. 20591

Controller Teamwork Evaluation and Assessment Methodology: A Scenario Calibration Study

Larry L. Bailey
Dana M. Broach
Richard C. Thompson
Robert J. Enos

Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

October 1999

Final Report

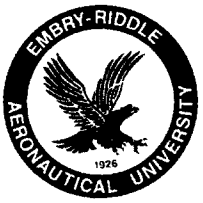
This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

U.S. Department of Transportation
Federal Aviation Administration

Doc
FAA
AM
99
24
c.2

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof.



JACK R. HUNT MEMORIAL LIBRARY
DAYTONA BEACH, FLORIDA 904-226-6595

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-99/24		

4. Title and Subtitle	5. Report Date
Controller Teamwork Evaluation and Assessment Methodology: A Scenario Calibration Study	October 1999
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Bailey, L.L., Broach. D.M., Thompson, R.C., and Enos, R.J.	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, OK 73125	11. Contract or Grant No.

12. Sponsoring Agency name and Address	13. Type of Report and Period Covered
Office of Aviation Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	14. Sponsoring Agency Code

15. Supplemental Notes

Work was accomplished under approved subtask AM-B-98-HRR-172.

16. Abstract

A low cost air traffic control (ATC) multi-sector training platform was developed to simulate radar-based air traffic control tasks. The purpose of the training device was to provide a vehicle for delivering ATC training on teamwork. However, before training could be delivered it was first necessary to develop training scenarios that would place participants under a specific amount of work. The results of the scenario calibration study reported in this paper suggest that the three scenarios can be viewed as representing low, medium, and high workload conditions based on the performance of 31 four-person teams. Statistically significant performance differences were observed across all three scenarios based on the percentage of aircraft that reached their destination within the allotted time, the amount of aircraft delay, the number of safety errors, and participants' perceptions of their workload.

17. Key Words	18. Distribution Statement
Air Traffic Control, Training, Teamwork	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

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

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

Controller Teamwork Evaluation and Assessment Methodology: A Scenario Calibration Study

Air traffic control specialists (ATCSs) do not work alone but must interact with others to ensure the safe, orderly, and expeditious flow of air traffic. For example, at airports, the ground ATCS coordinates taxi instructions with the tower cab ATCS to ensure that taxing aircraft do not cross the paths of aircraft landing or taking off. Similarly, when transferring control of an aircraft from one sector to another, an en route ATCS will coordinate the exchange with both the pilot and the ATCS from the adjacent sector.

ATCSs, however, receive relatively little formal training in crew coordination. Instead, they are selected, trained, and evaluated predominantly on their competencies. The assumption seems to be that technically competent ATCSs will develop the requisite coordination skills through on-the-job experiences (Hartel & Hartel, 1995).

Although past practices have produced the world's safest and most efficient ATC system, recent reviews have suggested that a more structured approach to ATCS team training is necessary (Wickens, Mavor, & McGee, 1997; Federal Aviation Administration, 1996; Mayberry, Kropp, Kirk, Breitler, Wei, 1996). In particular, what is needed are training devices and activities that will enable ATC teams to practice and to receive feedback on their crew coordination skills, and how those skills affect important system outcome measures such as safety and efficiency (Prince, Chidester, Bowers, & Cannon-Bowers, 1992).

Addressing the above needs, the Federal Aviation Administration's (FAA) Civil Aeromedical Institute (CAMI) began a program of ATCS team training research, entitled Controller Teamwork Evaluation and Assessment Methodology (CTEAM)[1]. Within the CTEAM environment, a low cost, high psychological fidelity multi-sector ATC team training research platform was developed to simulate radar-based air traffic control tasks. Before team training research could be conducted, however, it was first necessary to develop training stimulus materials (i.e., scenarios) that would place participants under a predetermined amount of workload.

In this study, three scenarios were developed to vary in difficulty based on the number of aircraft presented over time (i.e. aircraft density). An examination was then made to determine scenario effects on: (1) system outcome measures (i.e., aircraft delay time, safety errors, and percentage of aircraft reaching their destination), (2) subjective workload, and (3) the practice effects associated with repeated scenario administrations.

The CTEAM research platform is based on an extension of the single sector air traffic scenario test (ATST) developed for use as part of an ATCS selection battery (Broach & Brecht-Clark, 1994). In the ATST, participants are in complete control over how they manage the traffic flow within their sector. In contrast, within the CTEAM environment, participants must coordinate their activities with other players as they negotiate the transfer of aircraft from one sector to the next. Figure 1 shows the interlocking pattern of the four sectors comprising the CTEAM environment.

Two studies conducted using the ATST scenarios formed the basis for the hypotheses guiding this study (Gilliland & Schlegel, 1992; Seamster, Redding, & Broach, 1996). Gilliland & Schlegel evaluated the difficulty of 16 scenarios on the basis of the total amount of aircraft delay time, the number of safety errors, and the percentage of aircraft that reached their destination within the allotted time. They concluded that the single most important factor determining scenario difficulty was the number of aircraft presented over time (i.e., aircraft density). Their research also suggested that system outcome scores degraded with increase aircraft density. Therefore, we hypothesized that in the CTEAM environment:

- **Hypothesis 1a.** Increasing the level of aircraft density will produce incremental decreases in the percentage of aircraft reaching their destination.
- **Hypothesis 1b.** Increasing the level of aircraft density will produce incremental increases in the amount of aircraft delay.

[1] The reader is referred to Appendix A for additional information on: (a) the need for research on ATC team training, (b) an overview of CAMI's CTEAM research program, and (c) the development of the multi-sector ATC team training device.

- **Hypothesis 1c.** Increasing the level of aircraft density will produce incremental increases in the number of safety errors.
- **Hypothesis 1d.** Increasing the level of aircraft density will produce corresponding increases in the perception of participants' workload.

In a second study, the cognitive strategies employed by ATST participants were examined by Seamster, Redding, & Broach (1996). Seamster et al. found that participants quickly developed sector management strategies (i.e., traffic flow strategies) that were relatively resistant to change unless formal training was provided. However, when sector management training was provided, statistically significant improvements were achieved for aircraft delay times, the percentage of aircraft that reached their destination, but not for safety error scores.

Since the training that participants receive in the CTEAM environment is targeted at minimizing aircraft delay times, one might hypothesize that there will be minimal practice effects associated with delay scores and the percentage of aircraft reaching their destination. However, given that participants will still be on the steep end of the learning curve when they perform under experimental conditions, we expect that additional learning will take place between the first and second administration of a given scenario. This is because repeated exposure to a given scenario provides a form of experiential learning that could lead to a more efficient use of the initial training strategy, as well as other strategies that participants acquire. Thus, we expect to see improvements in system effectiveness measures (i.e., delay scores, safety error scores, and percentage of aircraft reaching their destination) as articulated in the following hypotheses:

- **Hypothesis 2a.** The percentage of aircraft reaching their destination will increase upon repeated exposure to a given scenario.
- **Hypothesis 2b.** The amount of aircraft delay will decrease upon repeated exposure to a given scenario.
- **Hypothesis 2c.** The number of safety errors will decrease upon repeated exposure to a given scenario.
- **Hypothesis 2d.** Subjective workload will decrease upon repeated exposure to a given scenario.

Improvements in system effectiveness, however, are not expected to be consistently observed across all levels of scenario difficulty. Although the training received will involve exposure to varying degrees of aircraft densities, we expect that participants will first develop optimal sector management strategies for the low aircraft density condition, and then, consistent with the findings of Seamster et al. (1996), they will attempt to use those strategies to accommodate increasing levels of aircraft density. However, due to the inefficiency of those strategies for higher aircraft density conditions, we expect that the amount of improvement in system effectiveness will diminish with increasing levels of aircraft density in the following ways:

- **Hypothesis 3a.** The amount of improvement in the percentage of aircraft reaching their destination will diminish with increasing levels of aircraft density.
- **Hypothesis 3b.** The amount of improvement in aircraft delay time will diminish with increasing levels of aircraft density.
- **Hypothesis 3c.** The amount of improvement in safety errors will diminish with increasing levels of aircraft density.
- **Hypothesis 3d.** The amount of improvement in subjective workload will diminish with increasing levels of aircraft density.

METHOD

Participants

Thirty-one 4-person teams, comprised of high school graduates (63% male) between the ages of 18 and 30 (the same age range as ATC applicants), participated in a scenario calibration study, which compared team performance under three levels of aircraft density. Since laboratory research using actual air traffic control specialists is expensive, it was first necessary to run preliminary scenario development tests before conducting research with ATCSs. Participants were recruited through a local temporary help provider and received $6.50/hour for two consecutive four-hour per day sessions.

Equipment

Currently, the CTEAM research platform consists of five (four clients and a server) 80486/DX2 66 personal computers operating under Windows NT 3.51. A scenario generation tool enables experimenters to control: (1) airport positions and landing directions, (2) when and where new aircraft enter a given sector, (3) the route of flight, (4) and the length

of a given scenario. While operating, the server also functions as an experimenter monitoring station by simultaneously displaying sector activity in a scaled down version of all four sectors.

In the CTEAM environment, airspace configuration and traffic are controlled to assess how system effectiveness measures (operational errors, aircraft delay time, and proportion of aircraft reaching their destination) change as a function of controller team composition (knowledge, skills, and abilities), task-specific, and team-oriented behaviors. Task-specific behaviors include the issuance of aircraft heading, altitude, and speed changes. Team-oriented behaviors include communication exchanges and the timing of those exchanges that transpire between controllers as they negotiate the transfer of aircraft from one sector to the next. Aircraft commands and controller communications are initiated by a mouse activated communication panel. A computer replay file is generated, enabling later playback and data extraction of individual and team performance.

Figure 2 shows a black-and-white screen shot from one of four CTEAM computer work stations. The screen is divided into four main sections: (1) the center portion, which depicts the airspace (i.e., sector) under the control of a given team member as well as a 10-mile view into the two adjacent sectors; (2) a panel on the right that is used to issue aircraft commands; (3) a panel at the bottom that is used to communicate with adjacent sectors (modeled after the en route air traffic control communication protocol; Federal Aviation Administration, 1993); and (4) a panel at the left that displays sector information (e.g., airport landing directions) and system effectiveness scores, which include safety errors (err), aircraft delays (del), and percentage of aircraft reaching their destination (pct.).

Within Figure 2's airspace are: (1) aircraft (shown as directional arrows) and their associated data block (first line includes a three-digit numeric aircraft identifier, followed by a two digit alpha sector identifier; the second line includes speed, altitude, and route of flight information, respectively); (2) airports (displayed as a circle cut by two parallel lines representing the runway) with a lower-case alpha identifier; and (3) sector gates (displayed as breaks in the solid lines outlining the sector) with an upper-case alpha identifier. There are two kinds of sector gates: (1) those which leave the playing field (exit gates); and (2) those which connect to adjacent sectors (hand-off gates). Figure 1's exit gates are "A" and "B." Its hand-off gates are "C" and "L."

The rules for managing intra-sector traffic are: (1) minimum aircraft separation is five miles from sector boundaries and other aircraft at the same altitude, (2) aircraft must land parallel to the runway at speed slow and level 1 (the lowest altitude), and (3) aircraft must depart through exit gates at speed fast and level 4 (the highest altitude). The rules for managing inter-sector traffic are: (1) aircraft cannot enter adjacent sectors without first receiving permission; and (2) unless negotiated otherwise, aircraft must depart through hand-off gates at speed fast and level 4.

Stimulus Material

Previous research on a single sector version of the experimental platform emphasized the need to develop scenarios with well-defined characteristics based on the: (1) number of aircraft presented, (2) route of flight, and (3) timing of scenario events to create specified workload conditions (Gilliland, & Schlegel, 1992). Consequently, three, 28-minute (1680 seconds) scenarios were developed to create three levels of difficulty based on the average number of aircraft presented to the team over time (i.e., aircraft density).

In the low-, medium-, and high-density conditions, aircraft were presented so that, for the majority of the time each sector would be managing an average of 1.89, 2.73, and 3.58 aircraft, respectively, as shown in Figure 3. Aircraft were presented in three, 660-second waves. The duration of a wave was developed by measuring the amount of time a novice pre-experimental team took to land the first aircraft under low-density conditions. In the last wave, the route of flight of aircraft was adjusted so that, theoretically, all aircraft could reach their destinations within the allotted time.

Aircraft originally appeared in an inactive state (except for one active aircraft at time zero) and were activated at the discretion of the originating sector. The location and timing of the presentation of inactive aircraft were the same in each sector. Once activated, the aircraft had to travel through the hand-off gates of three sectors before landing at the fourth (no exit gates were used in these scenarios). Thus, all team members were involved in the process of aircraft arriving at their destination.

Aircraft were introduced into a given scenario in 30-second intervals until the number presented matched a given scenario's required density level. This level was maintained until the projected landing time of the first aircraft at time 630 seconds. This was followed 30 seconds later with the entry of the first aircraft of the second wave (at time 660 seconds). Subsequent landings of first-wave aircraft appeared in 30-second intervals, beginning from time 630. Thus, in the high-density condition, an aircraft was introduced at time 0, 30, 60 and 90 seconds with projected landing times of 630, 660, 690, and 720 seconds, respectively. Second-wave aircraft were introduced at times 660, 690, 720, and 750 seconds. Because there was a 30-second delay between the landing of the first aircraft and the start of the second wave, the aircraft density level dropped by one for the duration of the transition interval, as shown in Figure 3.

Measures

Subjective workload was assessed using a variation of the National Aeronautic and Space Administration Task Load Index (TLX; Hart & Staveland, 1988). In the TLX, subjective workload is viewed as a multidimensional construct involving the subjective appraisal of one's: (1) mental demand, (2) physical demand, (3) temporal demand, (4) performance, (5) effort, and (6) frustration level. These dimensions were defined and presented as single items in a questionnaire format. Participants used a 19-point scale (1 = low, 19 = high) to indicate their responses. These were then summed to produce an overall measure of subjective workload.

Aircraft delay was computed in minutes based on the aggregation of activation delay and destination delay time. *Activation delay* represented the time it took for an aircraft to be activated after it appeared on the computer screen. *Destination delay* referred to the amount of time aircraft deviated from optimal flight paths (i.e., flying at the highest speed in the shortest line from point of aircraft entry to exit point).

Safety error scores were based on the aggregation of the number of separation errors (aircraft violating a 5-mile horizontal separation rule) and crashes that occurred.

Percentage of aircraft reaching destination was computed by dividing the number of aircraft that landed at their designated airports by the total number of aircraft eligible to land at a given airport. Since aircraft originate in one sector and fly through three sectors before landing at the fourth, this measure represents the best indicator of team performance. Higher percentages of aircraft landings indicate higher degrees of team coordination.

Training

During the first four-hour day, participants received part-task training on the use of the CTEAM function keys (as shown in Figure 1), and the aircraft separation and procedural rules used to manage intra- and inter-sector traffic. This was followed by three, 12-minute practice scenarios during which teams performed under each of three levels of aircraft density. Whereas the purpose of the former training was to develop individual sector management strategies, the purpose of the latter was to enable team members to adjust their individual strategies to accommodate the sector management strategies of adjacent sectors. Throughout the training, participants were taught to activate aircraft as quickly as possible. However, they were cautioned to only activate as many aircraft as they could safely manage.

Design and Procedures

During the second four-hour day, the three scenarios were administered in two time blocks for a total of six trials. This produced a 2 (time) x 3 (scenario difficulty) doubly-repeated measures experimental design. Within each time block, the order of scenario presentation was counterbalanced across teams to control for order effects. Participants were instructed to occupy the same sectors on which they had received training. This was to ensure that the coordination strategies that had been developed during the previous day's training would be applicable in the experimental setting. Participants had a 10-minute break after each trial with the exception of a 20-minute break between trials 3 and 4. Sector performance data were automatically recorded by the computer. This included the number of aircraft active across time. The data were aggregated at the team level of analysis by averaging across the four sectors. These data were then transformed into a time series by computing the mean number of active aircraft (across 31 teams) over a 28-minute period (1680 sec.) in 30-second intervals beginning at time zero. This produced a total of 57 sample points for each scenario.

RESULTS

Each scenario was designed to place participants under a predefined amount of work based on the average number of aircraft presented to the team over time. However, because aircraft were introduced in an inactive state, participants could control their workload by choosing to activate less than the designated number of aircraft, and thus compromise the intended level of scenario difficulty. That is, a high aircraft density scenario could be transformed into a low aircraft density scenario simply by team members choosing not to activate aircraft. Because of this, a manipulation check was first conducted to determine whether the introduction of inactive aircraft was sufficient motivation for participants to activate them as intended based, on the patterns shown in Figure 3.

Manipulation Check

Three questions guided the data analyses for the manipulation check. First, did participants activate the aircraft as planned? Second, did the average number of active aircraft create three statistically distinct scenarios? Third, were practice effects observed between the first and second administration of a given scenario?

Two analysis were conducted to determine whether participants activated aircraft as planned. In the first analysis, the expected mean aircraft densities (based on data from Figure 1) for the low-, medium-, and high-density conditions (1.89, 2.73, and 3.58, respectively) were used in one sample t-tests as the standard for comparison with actual aircraft densities. The results are shown in Table 1. No statistically significant differences ($p < .05$) were observed. Thus, for both time blocks, the average number of aircraft activated by participants corresponded to the expected values.

In the second analysis, a visual inspection of the scenario time series was used to determine how workload conditions (defined by aircraft density) varied as a function of time. Figures 4 and 5 show the time series pattern for actual aircraft densities in the first time block and second time blocks, respectively. In both cases, in contrast to the standard of Figure 3, there is an upward drift in the number of active aircraft. This suggests that participants were activating aircraft in successive waves before all aircraft had landed from the previous wave. Furthermore, there is a pronounced negative slope associated with each wave of actual aircraft densities. This suggests that aircraft were exiting the scenario throughout the duration of each wave either by landing or from crashes. Thus, in contrast to the constant work load conditions suggested by Figure 3 and the results of Table 1, data such as these illustrate that participants experienced varying workload conditions during the course of a given scenario.

To determine whether actual aircraft densities created three statistically distinct scenarios, the data were analyzed in two ways, using multiple comparison procedures as outlined by Toothaker (1993). First, temporal stability between scenario administrations was determined using pairwise t-tests. Then, within a given time block, the statistical significance of actual aircraft density differences across scenarios was determined by the Dunn method (Toothaker, 1993). This method adjusts the t-statistic to control for type 1 error associated with making multiple comparisons with the same data point.

All aircraft density differences across scenarios were statistically significant at $p < .01$. However, aircraft density conditions produced differential practice effects as shown in Figure 6 and Table 2. In the low-density condition, the mean number of active aircraft declined by 8% from the first to the second administration ($t = 5.03$, $df = 56$, $p < .01$); compared with a 2% increase for the high-density condition ($t = -3.21$, $df = 56$, $p < .01$). There were no statistically significant practice effects observed in the medium-density condition. These results indicate that the scenarios differed from one another, were relatively stable between first and second administrations, and consequently, were suitable for use as stimulus material.

Hypotheses Testing

A doubly repeated multivariate analysis of variance (MANOVA) was used to examine the effects that aircraft density had on a linear composite of subjective workload and the system outcome measures. Main effects were observed for time ($F = 30.84$, $df = 4$, $p < .01$) and scenario difficulty ($F = 49.71$, $df = 8$, $p < .01$), as well as a significant time by difficulty interaction ($F = 2.87$, $df = 8$, $p < .01$). Subsequent univariate analysis of the four dependent measures produced differential results as shown in Table 3. These results were further analyzed using the multiple comparison procedures previously described in Toothaker (1993). All within time block comparisons were statistically significant at $p < .01$, thus

providing support for Hypotheses 1a, 1b, 1c, and 1d. However, as expected, aircraft density conditions produced differential practice effects, as shown in Table 4.

Subjective workload. Main effects were observed for scenario difficulty ($F = 108.81$, $df = 2$, $p < .01$), and time ($F = 20.26$, $df = 1$, $p < .01$), as well as a significant time by difficulty interaction ($F = 5.73$, $df = 2$, $p < .01$). As Figure 7 and Table 4 show, the interaction was due to the differential drop in subjective workload between the first and second administrations of a given scenario. Under low, medium, and high aircraft density conditions, participants perceived a 16%, 8%, and 3% drop in workload respectively. With the exception of the high-density condition, these drops were statistically significant, thus providing support for hypothesis 3d and partial support for hypothesis 2b.

Percentage of aircraft reaching their destination. A main effect was observed for scenario difficulty ($F = 87.78$, $df = 2$, $p < .01$) and time ($F = 81.64$, $df = 1$, $p < .01$), as well as a significant time by difficulty interaction ($F = 3.98$, $df = 2$, $p < .05$). As Figure 8 and Table 4 demonstrate, the interaction was due to the practice effects associated with a differential increase in the percentage of aircraft that reached their destinations. Under low, medium, and high aircraft density conditions participants achieved respective increases of 14%, 16%, and 8% in the percentage of aircraft that reached their destinations. All increases were statistically significant, thus providing support for hypothesis 2a. However, only partial support was provided for hypothesis 3a. The amount of improvement in the percentage of aircraft reaching their destination did not uniformly decline with increasing aircraft density.

Aircraft delay time. Main effects were observed for scenario difficulty ($F = 348.58$, $df = 2$, $p < .01$) and time ($F = 44.42$, $df = 1$, $p < .01$). Under low, medium, and high-density conditions, participants experienced a 32%, 28%, and 16% drop in aircraft delay scores respectively. As Figure 9 and Table 4 show these drops were all statistically significant, thus providing support for hypotheses 2b and 3b.

Safety errors. In contrast to the previous three measures, a main effect was observed only for scenario difficulty ($F = 90.55$, $df = 2$, $p < .01$). These results failed to support hypotheses 2c and 3c. However, as Figure 10 and Table 4 show, within the low aircraft density condition there was a statistically significant 25% improvement ($p < .05$) in safety error scores. This suggests that participants' overall sector management strategies were optimal for the low aircraft density condition.

A summary of the overall results of the hypotheses tested appears in Table 5.

CONCLUSIONS

The results of this study demonstrated that CTEAM scenarios could be developed to produce standardized training conditions. Furthermore, through experiential learning, participants were able to improve their team coordination skills between the first and second administration of a given scenario.

CTEAM scenarios were shown to represent three statistically distinct levels of scenario difficulty (low, medium, and high) based on the incremental adjustment in aircraft densities. As scenario aircraft density increased there were statistically significant increases in the amount of aircraft delay times and the number of safety errors, and a decrease in the percentage of aircraft reaching their destinations. In addition to the objective performance indicators, subjective perceptions of workload also increased with increasing scenario difficulty.

Paralleling these objective performance indicators were the participants' subjective perceptions of their workload. These results indicated that scenario difficulty could easily be manipulated through the incremental adjustment in the number of aircraft presented over time. However, the results also demonstrated that scenario performance did not produce constant workload conditions, but instead, created a general pattern of increasing aircraft density throughout the course of a given scenario. Thus, CTEAM scenarios should be viewed as dynamic rather than static taskload environments.

Improvements in teamwork were also documented in this study by the increase in the percentage of aircraft that reached their destination between the first and second administration of a given scenario. Within each scenario, aircraft originated in one sector and flew through the remaining sectors before landing at the fourth. Thus, for aircraft to reach their destination, team members had to develop strategies not only for managing the traffic within their own sector, but also for coordinating the transition of aircraft between sectors. Since no additional training was provided between scenario administrations, im-

provements in teamwork were achieved solely through experiential learning, a desired attribute for a team training exercise.

Overall, the results of this study demonstrated that scenario performance was sensitive to the incremental adjustment in aircraft density. What remains to be demonstrated is whether scenario performance is sensitive to various training interventions targeted at improving ATC team coordination. Future research will more closely examine this issue by determining the extent to which various forms of team training produce improvements in CTEAM scenario performance above those achieved simply through experiential learning.

REFERENCES

Alexander, J., Ammerman, H., Fairhurst, W., Hostedler, C., & Jones, G. (1987). FAA air traffic control operations concepts, vol. Vi: ARTCC-host en route controllers (Final report delivered to the Federal Aviation Administration under contract DTFA-03-89-C-00023), Colorado Springs: CO: CTA Incorporated.

Broach, D. & Brecht-Clark, J. (1994). *Validation of the Federal Aviation Administration air traffic control specialist pre-training screen.* (DOT/ FAA/ AM-94/4). Washington, DC: Department of Transportation/Federal Aviation Administration. Available from: National Technical Information Service, Springfield, VA 22161; order #ADA277549.

Broach, D. (1992). Non-cognitive predictors of performance in radar-based air traffic control training. Paper presented at the Fourth Annual Convention of the American Psychological Society. San Diego, CA.

Collins, W., Nye, L.,& Manning, C. (1990). *Studies of poststrike air traffic control specialist trainees: III. Changes in demographic characteristics of Academy entrants and biomedical predictors of success in air traffic controller selection and Academy screening.* (DOT/FAA/AM-90/4). Washington, DC: Department of Transportation/Federal Aviation Administration. Available from: National Technical Information Service, Springfield, VA 22161; order #ADA223480.

Collins, W., Schroeder, D., Nye, L. (1989). *Relationships of anxiety scores to Academy and filed training performance of air traffic control specialists* (DOT/ FAA/AM-89/7). Washington, DC: Department of Transportation/Federal Aviation Administration. Available from: National Technical Information Service, Springfield, VA 22161; order #ADA209326.

Driskell,, J., Hogan, R., & Salas, E. (1987). Personality and group performance. *Review of Personality and Social Psychology, 9,* 91-112.

Federal Aviation Administration (1993). *Air traffic procedures handbook* (FAA Order 7110.65H). Washington DC: FAA Air Traffic Services.

Federal Aviation Administration (1996). *National airspace system architecture, version 2.0.* Washington, DC. Office of Systems Architecture and Program Evaluation (ASD).

Gilliland, K. & Schlegel, E. (1992). *Evaluation of extended practice effects on the air traffic scenarios test. Final report.* (Final report delivered to the Federal Aviation Administration under contract DTFA-02-92P13359). Norman, OK: University of Oklahoma.

Hart, S. & Staveland, E. (1988). Development of a NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Handcock & N. Meshkati (Eds.), *Human mental workload.* Amsterdam: North-Holland. 139-83.

Hartel, C., & Hartel, G. (1995). *Controller resource management - What can we learn from aircrews?* (DOT/ FAA/AM-95/21). Washington, DC: Department of Transportation/Federal Aviation Administration. Available from: National Technical Information Service, Springfield, VA 22161; order #ADA297386.

Helmreich, R. (1987). Exploring flightcrew behavior. *Social Behavior, 2,* 63-72.

Hodges, H. (1995). *Air traffic control sector resource management: A companion human factors concept to ensure the success of a data link-based air traffic control system.* Report to the Federal Aviation Administration. Crown Communications Incorporated.

Hogan, J. & Hogan, R. (1989). Noncognitive predictors of performance during explosive ordinance disposal training. *Military Psychology, 1*, 117-33.

Kahan, J., Webb, N., Shavelson, R., & Stolzenberg, R. (1985). *Individual characteristics and unit performance* (Rep. No. R-3194-MIL). Santa Monica, CA: Rand Corporation.

Mayberry, P., Kropp, K., Kirk, K., Breiter, A., & Wei, M. (1995). *Analysis of operational errors for air route traffic control centers: Final report.* (Final report delivered to the Federal Aviation Administration under contract DTFA01-95—C-00002). Alexandria, VA: CAN Corporation. Available from the Institute for Public Research Document Control and Distribution Section; order IPR95-5108.

Mogan, B., & Lassiter, D. (1992). Team composition and staffing. In R. Swezey & E. Salas (Eds.), *Teams: Their training performance.* Norwood, NJ: Ablex, 75-100.

Nye, L., & Collins, W. (1991). *Some personality characteristics of air traffic control specialist trainees: Interactions of personality and aptitude test scores with FAA Academy success and career expectations* (DOT/FAA/AM-91/8). Washington, DC: Department of Transportation/Federal Aviation Administration. Available from: National Technical Information Service, Springfield, VA 22161; order #ADA238027.

OU Vortec, Edwards, W., Fuller, D., & Manning, C. (1994). *Automation and cognition in air traffic control: An empirical investigation.* (DOT/FAA/AM-94/3). Washington, DC: Department of Transportation/Federal Aviation Administration. Available from: National Technical Information Service, Springfield, VA 22161; order #ADA277057.

Prince, C., Childester, T., Bowers, C., & Cannon-Bowers, J. (1992). Aircrew coordination - Achieving teamwork in the cockpit. In R. Swezey & E. Salas (Eds.), *Teams Their training performance.* Norwood, NJ: Ablex, 329-54.

Rodgers, M. & Nye, L. (1994). Factors associated with the severity of operational errors at air route traffic control centers. In M. Rodgers (Ed.), *An examination of the operational database for air route traffic control centers*, pp.11-25 (DOT/FAA/AM-93/22). Washington, DC: Department of Transportation/Federal Aviation Administration. Available from: National Technical Information Service, Springfield, VA 22161; order #ADA275986.

Schroeder, D., Broach, D. & Young, W. (1993). *Contribution of personality to the prediction of success in initial air traffic control specialist training* (DOT/FAA/AM-93/4). Washington, DC: Department of Transportation/Federal Aviation Administration. Available from: National Technical Information Service, Springfield, VA 22161; order #ADA264699.

Seamster, T., Redding, R., & Broach, D. (1996). Effects of performance strategies on the air traffic control simulation test. *Proceedings of the Human Factors & Ergonomics Society 40th Annual Meeting*, 1092-96.

Smith-Jentsch, K, & Baker, D. (1995). *Identifying tower cab team training needs: A conceptual framework and research plan.* (Draft report delivered to the Federal Aviation Administration under interagency agreement DTFA01-94-Z-02050).

Toothaker, L. E. (1993). *Multiple comparison procedures.* Newbury Park, CA: Sage.

Wickens, C., Mavor, A., & McGee, J. (Eds). (1997). *Flight to the future: Human factors in air traffic control.* Washington, D.C: National Academy Press.

FIGURES

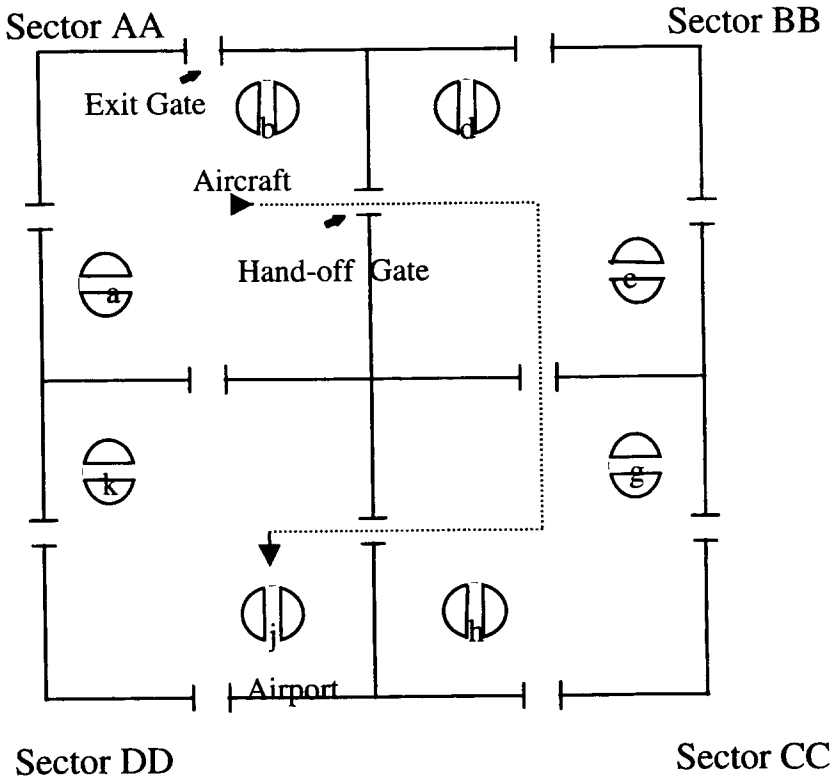


Figure 1. Sector configuration showing airport layout (lower case alpha), exit gates (at the perimeter), hand-off gates (at the interior), aircraft (arrow), and route of flight (dashed line).

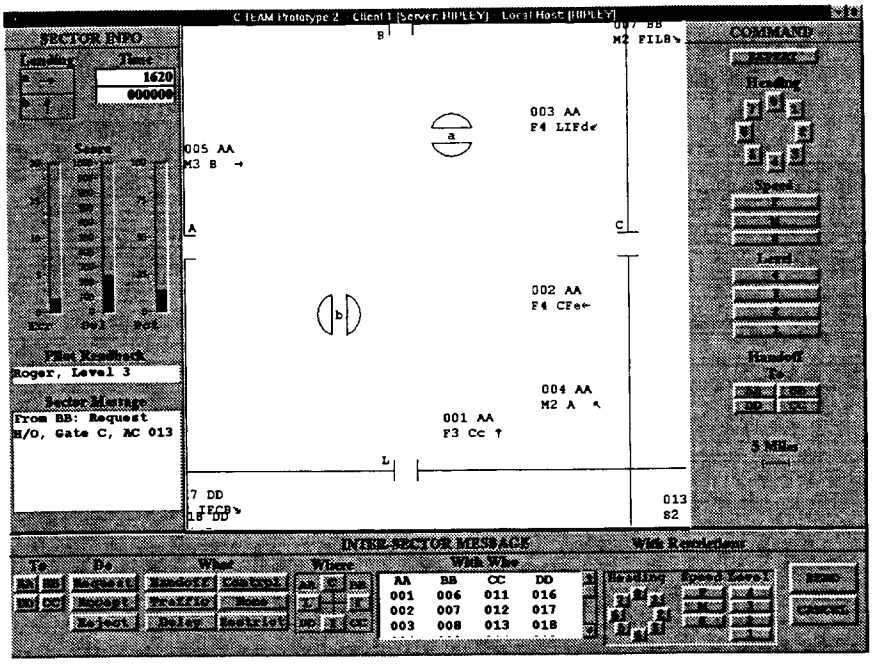


Figure 2. Screen shot from a CTEAM workstation showing sector airspace (middle), sector information (top left), aircraft commands (top right), and the controller-to-controller communication panel (bottom).

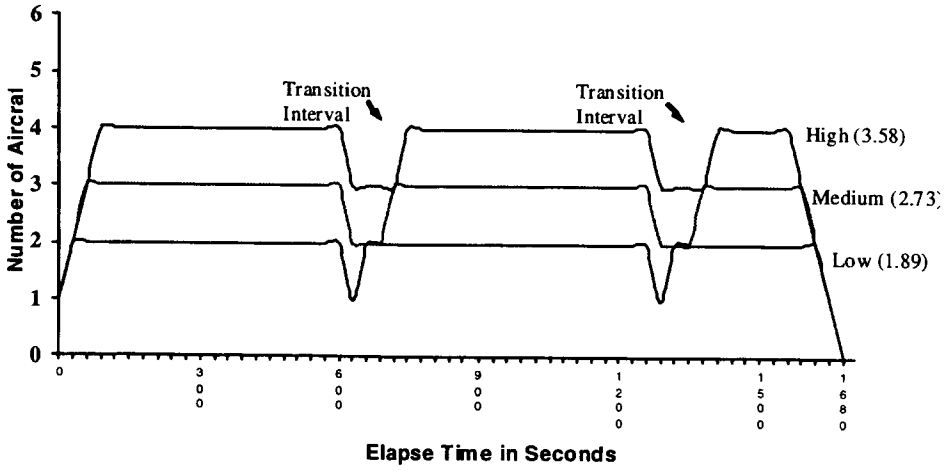


Figure 3. Standard aircraft density conditions across time for three levels of scenario difficultv.

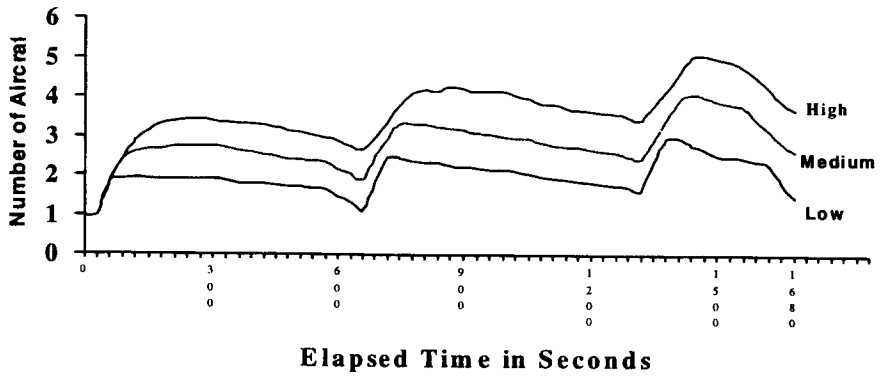


Figure 4. First time block: Average number of active aircraft for low, medium, and high aircraft density conditions.

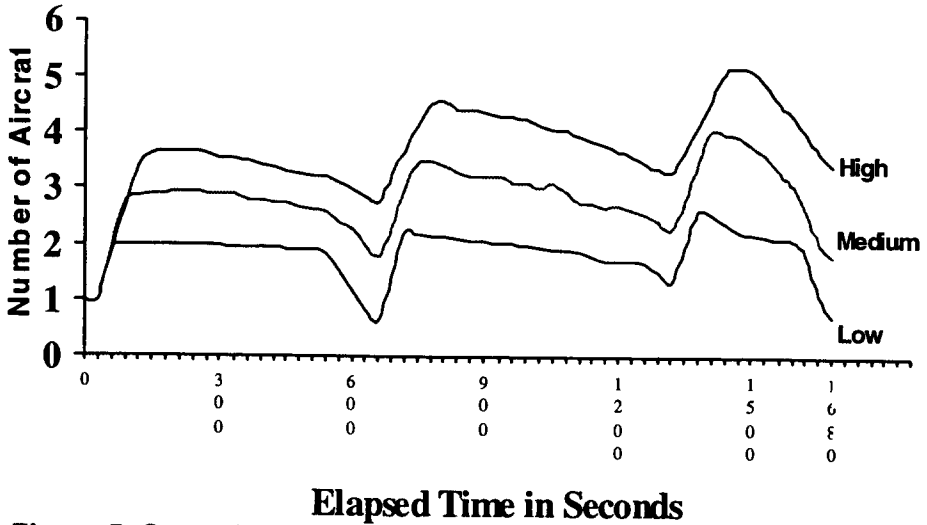


Figure 5. Second time block: Average number of active aircraft for low, medium, and high aircraft density conditions.

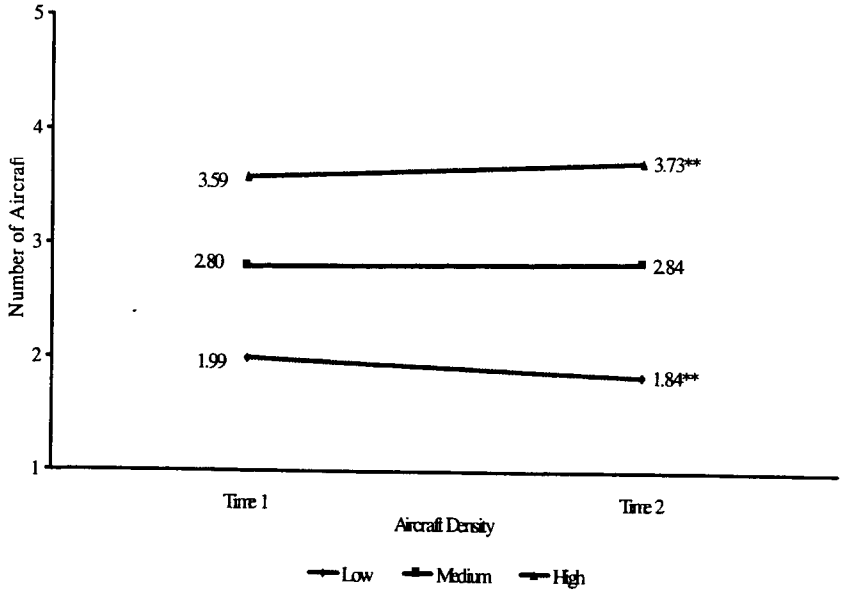


Figure 6. Actual aircraft densities. All differences between levels of aircraft density are significant at $p < .01$. Statistical significance between time block comparisons are indicated by ** = $p < .01$.

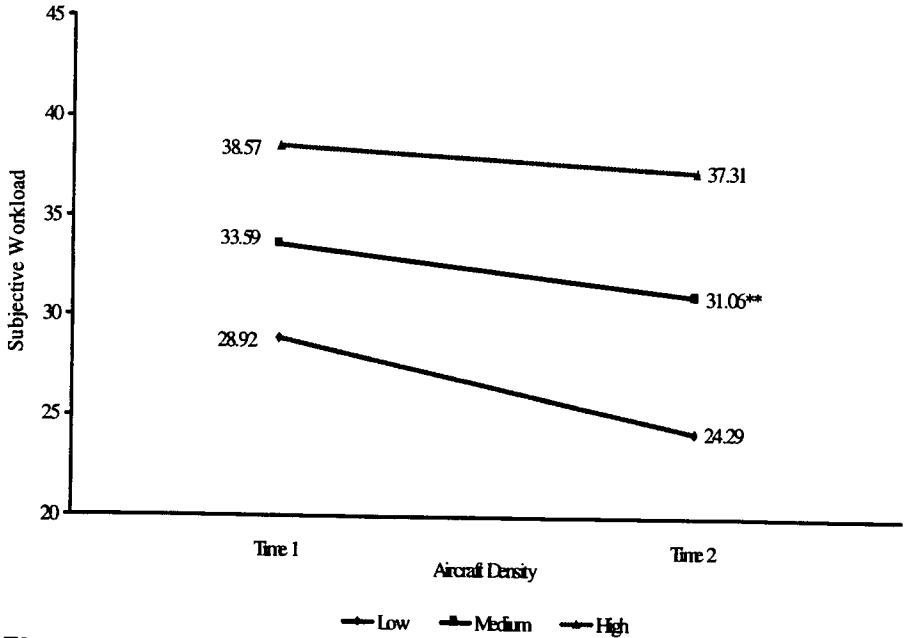


Figure 7. Subjective Workload. All differences between levels of aircraft density are significant at $p < .01$. Statistical significance between time block comparisons are indicated by ** = $p < .01$.

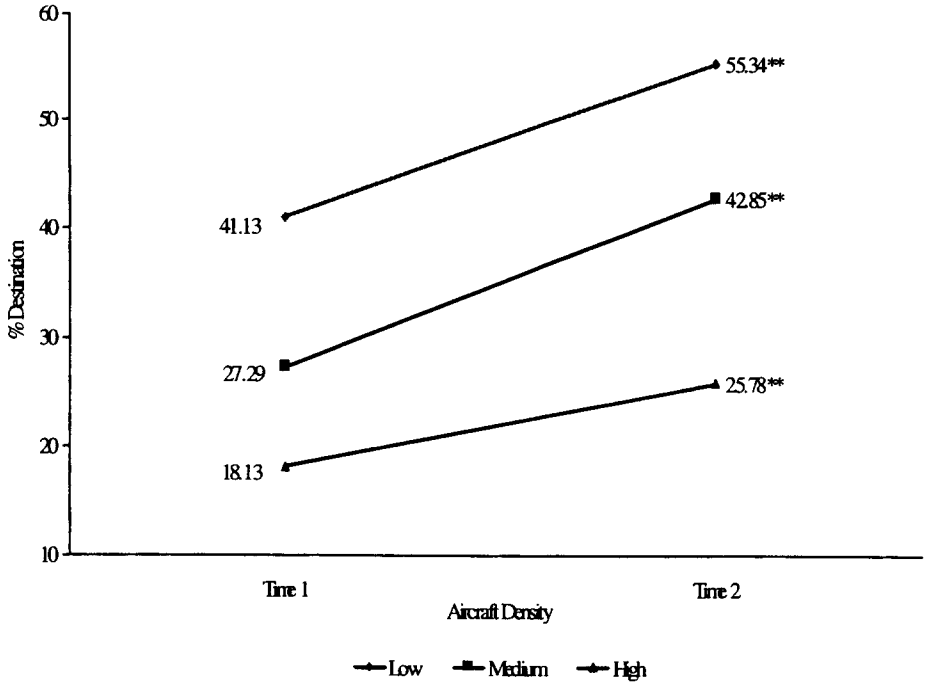


Figure 8. Percent of Aircraft Reaching Destination. All differences between levels of aircraft density are significant at $p < .01$. Statistical significance between time block comparisons are indicated by ** = $p < .01$.

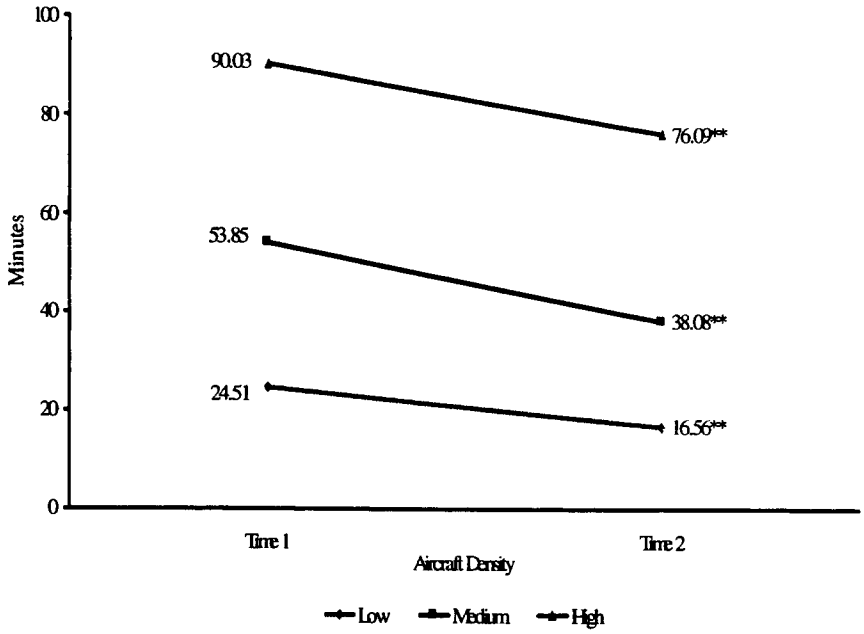


Figure 9. Aircraft Delay. All differences between levels of aircraft density are significant at $p < .01$. Statistical significance between time block comparisons are indicated by ** = $p < .01$.

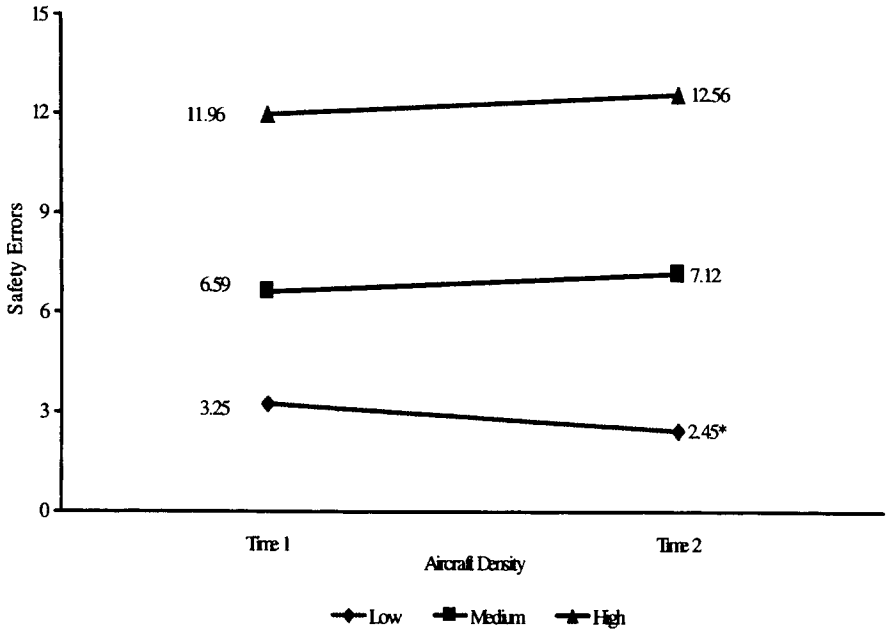


Figure 10. Safety Errors. All differences between levels of aircraft density are significant at $p < .01$. Statistical significance between time block comparisons are indicated by * = $p < .05$.

TABLES

Table 1. Actual Aircraft Densities Compared to Standard

	First Time Block					Second Time Block				
Scenario	Mean	SD	T	df	p	Mean	SD	t	df	p
Low	1.99	.42	1.76	56	.08	1.84	.42	-.97	56	.34
Medium	2.80	.61	.88	56	.38	2.81	.61	.99	56	.33
High	3.59	.83	.06	56	.95	3.68	.81	.97	56	.34

Table 2. Paired Comparisons of Actual Aircraft Densities

Scenario	t	df	p
Low	5.03	56	.00
Medium	-.27	56	.79
High	-4.28	56	.00

Table 3. Univariate tests for subjective workload, aircraft delay, safety errors, and the percentage of aircraft that reached their destination

Source	Measure	Type III Sum of Squares	df	F	p
Time	TLX	401.842	1	20.255	.000
	Delay	6715.868	1	44.419	.000
	Safety	.043	1	0.002	.965
	%Dest	.717	1	81.636	.000
Residual Error (Time)	TLX	555.496	28		
	Delay	4233.403	28		
	Safety	608.071	28		
	%Dest	.246	28		
Difficulty	TLX	3695.118	2	108.811	.000
	Delay	109973.381	2	348.581	.000
	Safety	2560.199	2	90.549	.000
	%Dest	1.912	2	87.780	.000
Residual Error (Difficulty)	TLX	950.851	56		
	Delay	8833.681	56		
	Safety	791.676	56		
	%Dest	.610	56		
Time by Difficulty	TLX	94.640	2	5.725	.005
	Delay	380.268	2	1.469	.239
	Safety	19.341	2	0.773	.466
	%Dest	.059	2	3.982	.024
Residual Error (Time by Difficulty)	TLX	462.890	56		
	Delay	7249.461	56		
	Safety	700.200	56		
	%Dest	.406	56		

Table 4. Time block paired comparisons of subjective workload (TLX), aircraft delay, safety errors, and percentage of aircraft reaching their destination

		First Time Block		Second Time Block		Paired Samples Test		
Scenario	Measure	Mean	SD	Mean	SD	t	Df	p
Low	TLX	28.92	4.39	24.31	5.53	4.77	30	.00
	Delay	24.51	10.99	16.56	10.34	6.31	30	.00
	Safety	3.25	2.92	2.45	2.32	2.58	30	.02
	%Dest	41%	20%	55%	18%	-6.32	30	.00
Medium	TLX	33.59	4.99	30.84	6.37	3.25	29	.00
	Delay	53.85	18.57	38.08	17.66	4.54	29	.00
	Safety	6.59	4.17	7.12	6.67	-.40	29	.69
	%Dest	27%	17%	43%	16%	-5.65	29	.00
High	TLX	38.20	4.25	37.31	6.23	1.33	30	.19
	Delay	90.03	20.45	76.09	22.50	3.86	29	.00
	Safety	12.01	5.00	12.39	6.55	-.35	29	.73
	%Dest	18%	10%	26%	10%	-4.93	30	.00

Table 5. Summary results of hypotheses testing

Hypotheses	Confirmed
1a. Increasing the level of aircraft density will produce incremental decreases in the percentage of aircraft reaching their destination.	Yes
1b. Increasing the level of aircraft density will produce incremental increases in the amount of aircraft delay.	Yes
1c. Increasing the level of aircraft density will produce incremental increases in the number of safety errors.	Yes
1d. Increasing the level of aircraft density will produce corresponding increases in the perceptions of participants' workload.	Yes
2a. The percentage of aircraft reaching their destination will increase upon repeated exposure to a given scenario.	Yes
2b. The amount of aircraft delay will decrease upon repeated exposure to a given scenario.	Yes
2c. The number of safety errors will decrease upon repeated exposure to a given scenario.	No
2d. Subjective workload will decrease upon repeated exposure to a given scenario.	Partially
3a. The amount of improvement in the percentage of aircraft reaching their destination will diminish with increasing levels of aircraft density.	Partially
3b. The amount of improvement in aircraft delay time will diminish with increasing levels of aircraft density.	Yes
3c. The amount of improvement in safety errors will diminish with increasing levels of aircraft density.	No
3d. The amount of improvement in subjective workload will diminish with increasing levels of aircraft density.	Yes

APPENDIX A.
CTEAM BACKGROUND INFORMATION

Need for Research on ATC Team Training. Sector resource management (SRM) refers to the available resources that air traffic controllers use to manage the traffic flow within a given airspace (Hodges, 1994). Examples of sector resources include air traffic control (ATC) procedures, radar displays, flight strips, auxiliary flight information displays, inter- and intra-facility communications, and communications with aircraft pilots. In addition, coordination with other controllers is required to manage traffic flow within and across sector boundaries. This coordination becomes especially important when there are deviations from the routine, such as disruptive weather, emergencies, and traffic flow restrictions.

Lapses in operational human performance, such as coordination, decision-making, and planning, have been specifically identified as factors underlying accidents and incidents in the National Airspace System (NAS). Specifically, coordination between controllers was cited as a causal factor in 15% of 1,038 low- to moderate-severity operational errors between 1988 and 1991 (Rodgers & Nye, 1994). More recently, the National Academy of Science noted a need to investigate controller teamwork in its review of human factors in air traffic control (ATC; Wickens, Mavor, & McGee, 1997).

Observationally, "teamwork" is comprised of the interactions between controllers, and controllers and NAS users. For example, the radar controller interacts with the assistant controller in identifiable patterns or sequences (O.U. Vortac, Edwards, Fuller, & Manning, 1994). Tasks requiring coordination and communication with other controllers, facilities, and NAS users were identified through job-task analysis of the ARTCC controller job by Alexander and his associates (Alexander, Ammerman, Fairhurst. Hostedler, & Jones, 1987).

Each controller brings to this interactive, real-time environment of the NAS certain characteristics that can influence the effectiveness and efficiency of his or her interactions with other controllers and NAS users. These characteristics include (a) general and task-relevant abilities, (b) ATC-specific knowledges and skills, (c) biographical experiences, and (d) personality. A baseline description of the relationship between general and task-relevant abilities and individual performance outcomes was developed in FY97 by the Air Traffic Selection and Training (AT-SAT) program. However, the program did not address the measurement or prediction of teamwork. Similarly, substantial work has been done on the relationships between ATC-specific knowledges and skills and ATC task performance at the individual level through programs such as the controller cognitive task analyses and enroute curriculum redesign. However, less research has been conducted with the measurement of teamwork as its focus. Biographical data have demonstrated some validity in the prediction of ATC training outcomes (Broach, 1992; Collins, Nye, & Manning, 1990). In other settings, biographical factors such as age, sex, ethnicity, socioeconomic status, and educational level have been shown to affect teamwork, largely through experimental manipulation of team composition in experimental settings (Morgan & Lassiter, 1992). But no experimental studies of ATC team composition have been conducted to date. Finally, there is considerable debate about the relationship of personality to teamwork (Morgan & Lassiter, 1992). On one hand, some researchers claim that prior findings have been inconclusive, providing a weak case at best (Kahan, Webb, Shavelson, & Stolzenberg, 1985). On the other hand, some scientists report that personality characteristics of individual team members can have a significant effect on team performance (Driskell, Salas, & Hogan, 1987; Helmreich, 1987; Hogan & Hogan, 1989). Previous research on controllers found that personality traits were useful predictors of training outcomes at the individual level of analysis (Collins, Schroeder, & Nye, 1989; Nye & Collins, 1991; Schroeder, Broach, & Young, 1993). But the relationship of

personality to teamwork, that is, the efficiency and effectiveness of interactions between controllers and the outcomes of those interactions in terms of safety and efficiency, has not been investigated.

Overview of CTEAM Research Program. In 1994, the Federal Aviation Administration (FAA) initiated a research program on controller teamwork in cooperation with the Naval Air Warfare Center Training Systems Division. The Navy's effort was directed at addressing team training issues in the tower cab environment (Smith-Jentsch & Baker, 1997). Work at the FAA's Civil Aeromedical Institute (CAMI) focused its efforts on team training research in the en route environment. CAMI's research program was entitled Air Traffic Controller Team Evaluation and Assessment Methodology (CTEAM). Its team training research agenda was to: (1) develop an experiential, high psychological fidelity, low-cost, team-based simulation of radar air traffic control tasks; (2) develop computerized scenario- generating capabilities; (3) design and calibrate scenarios to give CTEAM participants a specified amount of workload while addressing field-related inter-sector coordination situations; (4) develop a behavioral model of teamwork within the context of the CTEAM simulation; (5) test and validate diagnostic team assessment and feedback tools suitable for team training; (6) assess the effects of different team training interventions on objectively measured team and individual performance; and (7) investigate the potential of the PC-based simulation as a performance-based selection tool by evaluating the effects of team composition on safety and efficiency as well as its relationship to existing, planned, and other potential selection tools.

Development of CTEAM software. CTEAM was developed using a rapid-prototype spiraling requirements methodology. This approach allowed the researchers to review intermediate versions and to suggest software refinements based upon experiences with operating the system. The initial prototype was little more than an interface based upon a researcher's drawing, while succeeding prototypes added functionality until a final system was agreed upon. The data, which were provided by the system for research and feedback, were also refined through this process. After the final CTEAM simulation was completed, additional software pieces, such as a fully featured scenario generator and post-run data extraction tool, were developed.

The system was initially developed using Microsoft Windows NT named pipes for communications. As time progressed, this choice was superseded by TCP/IP sockets. The reasons for this were two-fold: apparent bugs in the named pipe communications and the desire to run the system under Windows 95 (Windows 95 can only connect to named pipes, it cannot create them). The internal system architecture is still designed around named pipes and does not make the most efficient use of sockets.

The CTEAM server is multi-threaded with a new thread created to handle each client. When the system was first designed, it was expected it would run on moderate sized 486 systems with multiple processors. The multi-threading would allow the systems to take full advantage of these machines. However, with the advent of newer, faster, and cheaper processors, the multi-threading is not necessary. The clients are single threaded.

The CTEAM system was written using Visual C++ 2.0 and targeted for Windows NT 3.1. Many of the interface controls were developed specifically for use by the system. This was necessary to provide the visual effect desired by the researchers. As it exists, the system is quite functional; however, from a maintenance and enhancement standpoint, it could benefit from being rewritten using the code generator tools now available with Visual C++ 4.2/ 5.0 and the MFC libraries. A JAVA version would also have several advantages, particularly since the clients could be run from within a web browser over the Internet.