AD-A277 549

Validation of the Federal Aviation Administration Air Traffic Control Specialist Pre-Training Screen

**DTIC
ELECTE
MAR 2 9 1994
S F D**

Dana Broach
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

Jan Brecht-Clark
Federal Aviation Administration
Office of Aviation Medicine
Washington, DC 20591

February 1994

Final Report

94-09542

U.S. Department
of Transportation

**Federal Aviation
Administration**

9 4 3 28 100

Best
Available
Copy

NOTICE

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

Technical Documentation Page

1. Report No. DOT/FAA/AM-94/4	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Validation of the Federal Aviation Administration Air Traffic Control Specialist Pre-Training Screen		5. Report Date **February 1994**
		6. Performing Organization Code
7. Author(s) Dana Broach, Ph.D., and Jan Brecht-Clark, Ph.D.		8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, OK 73125		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S. W. Washington, DC 20591		13. Type of Report and Period Covered

15. Supplementary Notes
This research was conducted under task AAM-C-93-HRR-129 and AAM-C-HRR-163

16. Abstract

Two formal validation studies of the Air Traffic Control Specialist Pre-Training Screen (ATCS/PTS), a 5-day computer-administered test battery, are described. The ATCS/PTS was designed to replace the 9-week U.S. Federal Aviation Administration (FAA) Academy ATCS Nonradar Screen program that served as the second major test in the ATCS selection system. Review of ATCS job analyses suggested that predictor tests should assess cognitive constructs such as spatial reasoning and short-term memory, and require dynamic, concurrent performance. A proposed test battery was developed, consisting of 2 computer-administered information processing tests and a simplified radar-based air traffic control work sample. In study 1, predictive, criterion-related validation (N = 438) found that the proposed test battery explained additional variability in scores earned in the 9-week FAA Academy program, after taking into account student aptitude. In study 2, criterion-related validation (N = 297) demonstrated that the proposed test battery was as valid as the 9-week FAA Academy ATCS Nonradar Screen for predicting progress in field training. Preliminary data from a third study conducted after validation of the ATCS/PTS seem to suggest that the abilities assessed by the new computerized tests reflect the abilities required on the job. However, implementation of the ATCS/PTS for actual employment decisionmaking in June 1992 was based on results obtained in the second concurrent, criterion-related validation study. The U.S. controller selection system since June 1992 has consisted of the 4-hour written ATCS aptitude test battery followed by, for those applicants earning a qualifying score and dependent upon agency manpower requirements, second-level screening on the ATCS/PTS. Additional research requirements as part of an aviation human factors research program are described.

17. Key Words Air Traffic Control Specialist, Selection, Validation, Tests, Ability, Job Analysis, Computer-administered Test	18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, VA 22161		
19. Security Classif. (of this report) UNCLASSIFIED	20. Security Classif. (of this page) UNCLASSIFIED	21. No. of Pages 16	22. Price

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

ACKNOWLEDGMENTS

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

VALIDATION OF THE FAA AIR TRAFFIC CONTROL SPECIALIST PRE-TRAINING SCREEN

The United States Federal Aviation Administration (FAA) is charged with managing the U.S. airspace. Air traffic controllers are at the heart of a web of radars, computers, and communication facilities that ensure the safety and efficiency of an increasingly complex air transportation system. Appropriate selection of personnel into a training program for these critical positions is an important human factors problem. This paper describes research conducted by the FAA to validate a cost-effective air traffic control specialist (ATCS) selection procedure. The project resulted in the implementation of a new selection test for ATCS applicants in June 1992 that was radically different from any previous ATCS selection program undertaken in the U.S.

Project background

The ATCS selection process between fiscal years 1986 and 1992 consisted of two major tests: (a) a 4 hour written aptitude examination administered by the United States Office of Personnel Management (OPM); and (b) a 9-week initial training program administered by the FAA Academy. Between 1984 and 1992, over 200,000 applicants took the written OPM aptitude examination across the country at a cost of about $20 per examinee (J. Aul, personal communication). Between October 1985 and January 1992, just 12,869 of those 200,000+ applicants were selected to attend the FAA Academy ATCS Nonradar Screen ("ATCS Screen"). The direct cost of this second-stage in the selection process was about $10-12,000 per student (Gwen Sawyer, June 1990). Of those students entering the ATCS Screen, 7,091 successfully graduated and entered into on-the-job training. This two-step selection process cost the FAA annually between $20 and 25 million to obtain approximately 1,400 trainee or "developmental" controllers.

The written aptitude tests - ATCS Screen selection process also imposed significant costs on applicants. Applicants selected to attend the ATCS Screen had to leave their current jobs and families for 9 weeks with a 55 - 60% chance of remaining in the controller occupation at the end of the program. That risk may have discouraged potentially qualified women and minority persons from pursuing an air traffic career (ASI, 1991). The FAA undertook a major review of its ATCS selection and training programs in 1990 to address these agency and applicant costs and other concerns. Three major ATCS selection policy goals were identified for the project: (1) reduce the costs of ATCS selection; (2) maintain the validity of the ATCS selection system; and (3) support agency cultural diversity goals. The first step toward achieving these goals was to develop and validate a test battery to replace the 9 week ATCS Screen.

Proposed test battery

Development of the new test battery began in late 1990 by reviewing available information about the cognitive requirements of the ATCS job. As described in one recent cognitive task analysis, controllers attend to multiple information sources, assess and integrate the data, develop and prioritize plans of action, and implement those plans under time pressure while maintaining situational awareness (Human Technology, Inc., 1991). To assess the cognitive and sensory attributes required to perform these job functions, a proposed test battery was developed by ASI. The battery was developed within the conceptual framework provided by Multiple Resources Theory (Rodriquez, Narayan, & O'Donnell, 1986; Shingledecker, 1984; Wickens, 1984). Two computer-administered information processing tests were designed to dynamically assess cognitive attributes such as spatial reasoning, short-term memory, movement detection, pattern recognition, and attention allocation (ASI, 1991). In addition, a low-fidelity radar simulation of air traffic control vectoring and separation tasks was also developed as a computer-administered work sample. The information processing tests and the work sample required performance of concurrent, multiple tasks by candidates to reflect the job demands placed on controllers.

The 2 computerized information processing tests were (a) the Static Vector/Continuous Memory test (SV/CM) and (b) the Time Wall/Pattern Recognition test (TW/PR). In the Static Vector (SV) component of the first test, a pair of simulated aircraft were presented on the left half of the computer screen (Figure 1). A quasi-data block for each target gave speed ("S250" was 250 knots), altitude ("A250" meant 25,000 feet). The subject's task was to determine as rapidly and accurately as possible if the simulated aircraft were in conflict based on their altitude, speed, and spatial relationship. The Continuous Memory (CM) component on the right side of the screen presented 2 aircraft call signs, one above and the other below a line. The subject's task was to remember the *bottom* call sign ("Target call sign" in Figure 1), for in the next CM trial, the subject had to indicate if the call sign *above* the line ("Probe call sign" in Figure 1) was the same as had been presented *below* the line in the previous CM trial. However, the subject had to encode what was

now the bottom call sign before responding, for as soon as an answer was made, a new set of call signs appeared. The attention director at the bottom center of the SV/CM screen informed the subject which task (SV or CM) was to be performed for each trial. A fixed number of trials for each component (SV and CM) were administered in a 5 minute SV/CM session. The speeds, altitudes, and spatial relationships between aircraft in the SV and the call signs in the CM varied from trial to trial within the session. Performance feedback was provided at the end of each session on each component (SV and CM).

The TW/PR test also consisted of a set of paired tasks (Figure 2). In the Time Wall (TW) component, a square target appeared first, moving from left to right at a steady speed toward the "wall" on the far right of the screen. After an initial time interval, the moving target and wall disappeared and were replaced by pairs of patterns. The Pattern Recognition (PR) task was to decide if the patterns were identical while keeping in
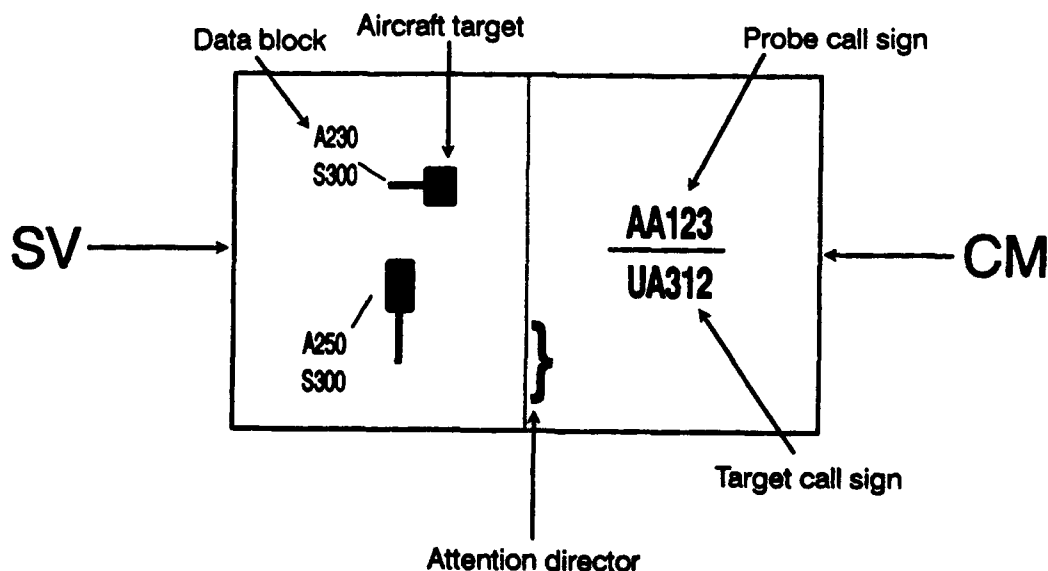

FIGURE 1. STATIC VECTOR (SV)/CONTINUOUS MEMORY (CM) SCREEN. SV test is shown on the left-hand side of the screen, CM test on the right. When the attention director was to the left, the subject's task was to decide if the aircraft targets would collide or not, based on the altitude ("A230") and speed ("S300") information in the data blocks and spatial relationships of the targets. When the attention director was to the right, the subject's task was to first, memorize the target call sign below the line, and second, indicate if the probe call sign above was the same, or different, as the target call sign that had been presented below the line in the previous CM trial.

mind the continuing movement of the TW target toward the wall. The TW task was to stop the now invisible target as close as possible to, without actually hitting or passing through, the wall. Subjects were presented with a fixed number of TW/PR trials within a nominal 5-minute test session; the actual length of the session was a function of subject response time. For example, consistently stopping the moving target in the TW short of the wall by a large margin reduced total session time proportionately. Measures from both the SV, CM, and PR components included the mean percent correct and mean reaction time for correct responses across trials within the 5-minute sessions for each test pair; the TW measure was the absolute distance (in milliseconds) between the wall and target when stopped by the subject. Performance feedback on these measures was provided to the subjects at the end of each 5 minute session.

The Air Traffic Scenario Test (ATST; Figure 3), the computer-administered work sample component of the proposed test battery, was developed by 4 subject matter experts with more than 30 years of air traffic control experience (ASI, 1991). The task required the subject to control aircraft within a simplified synthetic airspace, directing them to their destinations according to a small set of rules. There were 6 destinations: 4 outbound gates, A, B, C, and D; and 2 airports, E and F. The direction of travel, speed, and altitude of the aircraft, represented by small arrows next to the quasi-data blocks, were controlled by mouse. Three alphanumeric characters comprised the quasi-data blocks: first, aircraft speed (*Slow, Medium, Fast*); second, altitude (*1 = Lowest, 4 = Highest*); and third, destination. The orientation of the aircraft arrow indicated its current direction of flight. An open circle in an upper corner of the data block indicated an aircraft waiting to be
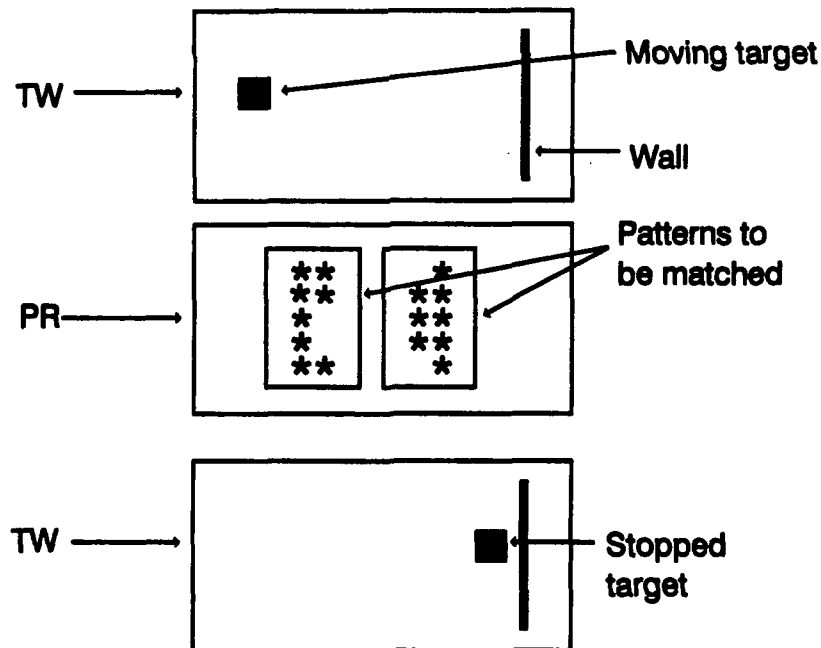
FIGURE 2. TIME WALL (TW)/PATTERN RECOGNITION (PR) SCREENS. *First, the target appeared, moving from left to right at a steady speed toward the "wall" (Top screen). After an initial time interval, the target and wall were masked by a pair of patterns (Middle screen). The subject's task was to decide if the patterns were the same or different. A new pair of patterns appeared after each response was made. However, the subject had to keep in mind the continuing movement of the TW target toward the wall, as the TW task was to stop the target (Bottom screen) as close as possible to, without actually hitting or passing through, the wall.*

3

activated by ("handed off to") the subject. The large arrow in the lower right hand corner of the screen indicated the landing direction at airports E and F, while the bottom horizontal bar icon represented the minimum lateral separation distance. Aircraft landed at airports E and F at the lowest altitude and slow speed in the required direction; aircraft exited gates A, B, C, and D at the fastest speed and highest altitude. A difference in altitude between any two aircraft was considered adequate separation; aircraft at the same altitude had to be separated by at least 5 nautical miles as represented by the separation icon. In addition, all aircraft had to be separated from the airspace boundary by at least 5 nautical miles. Error counts were obtained and summed to create an overall error score. In addition, the system automatically computed the difference between the actual time to reach destination for each aircraft and the time required for the optimum flight path as determined by the system software. This

en route delay time was summed with the time each aircraft spent waiting to be activated as a measure of overall controller efficiency. Performance feedback on these measures was provided to subjects at the end of each of 20 practice scenarios.

Study 1:
Predictive, Criterion-related validation

Two validation studies of this proposed test battery were conducted by the FAA in 1991. The purpose of the first study was to assess the predictive, criterion-related validity of the proposed test battery, and to determine the incremental validity of the proposed computerized tests over the existing written test. The sample in the first predictive, criterion-related validation study consisted of the 423 newly hired air traffic control students who entered the ATCS Screen in March and April 1991 in accordance with existing FAA procedures and policies. The sample was pre-
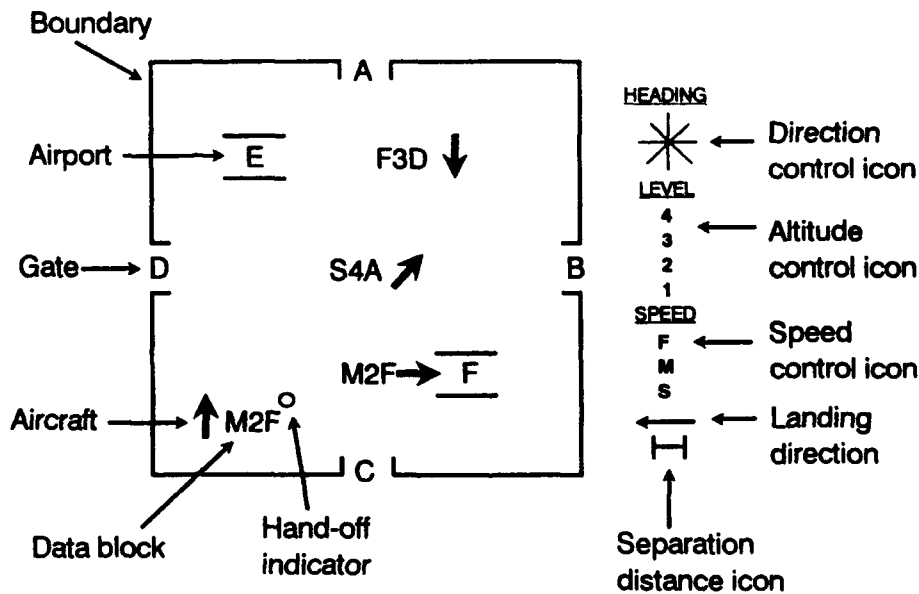

FIGURE 3. *AIR TRAFFIC SCENARIO TEST (ATST) SCREEN. The boundary encloses a simplified airspace, with 4 outbound gates, A, B, C, and D and 2 airports, E and F. The aircraft and direction of flight are represented by the arrows adjacent to a data block. The alphanumeric data block indicates aircraft speed (S, M, or F) and altitude (1 = lowest, 4 = highest). Aircraft waiting to be handed off are tagged with a small open circle in the upper right hand corner of the data block. Aircraft are controlled with a mouse. First, the candidate clicks on an aircraft, and then clicks on the appropriate element of either the direction control, altitude control, or speed control icons to change that flight parameter. Subjects are reminded of the required landing direction at airports and minimum horizontal separation distance by the landing direction and separation distance icons respectively.*

4

dominantly male (77.1%) and non-minority (83.0%); most (88.7%) had entered federal service by competitive examination rather than by non-competitive special appointment. There were significantly more women (22.9%) in this validation sample compared to the population of students that had entered the ATCS Screen between October 1985 and January 1991 (18.9%; $Z = 2.05$, $p \leq .05$). Similarly, minorities were also over-represented in this validation sample (17.0%) in comparison to the population of ATCS Screen students (10.2%; $Z = 2.38$, $p \leq .05$). The majority (65%) had no prior aviation-related experience which was representative of the population that reached the 2nd stage of the ATCS selection system. Aptitude scores for the ATCS occupation, represented in this study by the variable *RATING*, are based on the civil service test scores earned by an applicant on the written aptitude test plus any statutory veteran's preference points. The general development, psychometric characteristics, and validity of the written aptitude test battery has been extensively described (Sells, Dailey, & Pickrel, 1984). *RATING* was used to rank-order competitive applicants within statutory guidelines such that hiring was done on the basis of merit (Aul, 1991).

Criterion for predictive validation

The criterion for this predictive study was the final composite score earned in the ATCS Screen. The ATCS Screen was originally established in response to recommendations by the U.S. Congress House Committee on Government Operations (U.S. Congress, 1976) to reduce field training attrition rates. The ATCS Screen was based upon a miniaturized training-testing-evaluation personnel selection model (Siegel, 1978, 1983). Thirteen performance assessments, including classroom tests, laboratory simulations of nonradar air traffic control, and a final written examination, were made during the course of the ATCS Screen (Della Rocco, Manning, & Wing, 1990). The final summed composite score (*SCREEN*) of these ATCS Screen performance measures was weighted 20% for classroom tests, 60% for laboratory scores, and 20% for the final examination, with a minimum score of 70 out of 100 required to pass. In this sample, 56.0% passed the ATCS Screen, 27.7% failed, and 16.3% withdrew prior to completion. The mean

SCREEN score of 71.8 ($SD = 11.8$) for this validation sample of 423 students was not significantly different from that of the population students that had entered the ATCS Screen between October 1985 and January 1991.

Procedure

The proposed test battery was administered in 2 waves to subjects the week prior to beginning the ATCS Screen. The subjects were tested in March and April 1991 at the FAA Civil Aeromedical Institute (CAMI) in Oklahoma City. Instructions on the test battery were given on Monday morning. A total of 20 SV/CM and 20 TW/PR practice sessions were administered to subjects across 3.5 days (Monday afternoon through Thursday). The SV/CM and TW/PR tests did not change in difficulty across sessions. Subjects also were given 20 practice scenarios for the ATST, building in complexity and difficulty from about 12 aircraft in 30 minutes to over 40 aircraft in less than 30 minutes in the final practice sessions. Performance feedback was provided to subjects after each practice session. On Friday, subjects received a final 4 SV/CM, 4 TW/PR sessions, and 6 ATST scenarios. Measures were averaged across these final graded sessions within test, yielding 8 proposed test scores: (1) SV average percent correct; (2) SV average correct response reaction time; (3) CM average percent correct; (4) CM average correct response reaction time; (5) TW average absolute error; (6) PR average correct response reaction time; (7) average ATST error score; and (8) summed delay and waiting times in the ATST scenario. Aptitude ratings and ATCS Screen scores were extracted for the 423 subjects from the CAMI research data bases after all subjects had completed the ATCS Screen. These data were matched with proposed test scores for analysis; proposed test scores were not used in any way to make employment decisions about the subjects.

Results

On one hand, performance on the SV/CM and TW/PR tests appeared to reach differential stability (Bittner, 1979) at about the 15th session. The average performance within test component across the final sessions represented a reasonable measure of asymptotic

individual differences on those tests. On the other hand, learning curve analyses were not possible with the ATST because scenario difficulty increased across sessions. However, average performance across the final 6 scenarios was still computed as the index of individual differences on that test component. Multiple regression analysis was used to assess how well the proposed test battery predicted student performance in the ATCS Screen after taking into account student aptitude. First, *RATING* was entered into the regression equation predicting *SCREEN*. There was a statistically significant linear relationship between *RATING* and *SCREEN* of $R = .23$, p .001, where R was the multiple correlation between predictor (*RATING*) and criterion *(SCREEN)* and $p \leq .001$ indicated that an R of this magnitude would be expected by chance alone in less than 1 in a thousand times. The relationships of proposed test battery average final scores to *SCREEN* were analyzed in the second step of the multiple regression analysis using a forward stepwise procedure to determine the optimal combination of predictor variables. In a forward stepwise multiple regression analysis, the proposed test score accounting for the most variability left in the criterion *SCREEN* entered the regression equation; then, one at a time, proposed test scores which accounted for the most of the remaining unexplained variability in *SCREEN* were added to the equation, until the amount of variability explained by a new score became insignificant. The optimal linear combination of proposed test scores accounted for an additional 20% ($R^2 = .20$, $p \leq .001$) of the variability in *SCREEN* over the proportion of variability already explained by student aptitude scores (*RATING*). There were no statistical differences in the prediction equation by sex and minority status (ASI, 1991), suggesting that the proposed test battery might not adversely impact protected classes of applicants.

Study 2:
Concurrent, criterion-related validation

Encouraged by the results of the initial predictive study, the FAA conducted a concurrent, criterion-related validation study to assess the validity of the proposed test battery as a replacement for the ATCS Nonradar Screen (Weltin, Broach, Goldbach, & O'Donnell, 1991). The sample for this second valida-

tion study was composed of 297 trainee ("developmental") and FPL controllers. While this sample was predominantly male (64.6%) and non-minority (61.6%), women and minorities were over sampled relative to their representation in the ATCS workforce. The majority of the sample was drawn from en route centers (58.2%), reflecting the historical employment patterns in the workforce; 49.2% had attained FPL certification. The final composite *SCREEN* score for each participant was extracted from the CAMI ATCS Selection data base and used as the current predictor in this study. The SV/CM, TW/PR, and ATST average test scores described in the first study were the alternative predictors in this validity study. The ATCS Pre-Training Screen (ATCS/PTS), as the proposed battery had come to be known, was administered to subjects during late summer 1991 using the same test administration protocols as in the first study.

Criterion for concurrent validation

This study was constrained to use available training performance indices as validation criteria; no other criteria were developed or collected. These indices included the number of days spent in particular phases of field training and hours of formal, documented on-the-job training (OJT) provided under the supervision of a designated OJT Instructor within those phases, as reported by field ATC facilities in accordance with national policy (FAA, 1985). Subjective ratings of developmental performance in that phase of training (1 = *Bottom 10% compared to all other controllers observed in training*, 6 = *Top 10% compared to all other controllers observed in training*) were also available for each participant in this second validation study. Data for the ground, local, and radar control phases of instruction were extracted from the CAMI ATCS Training Tracking data base for subjects drawn from FAA terminal facilities. The ground control phase qualified a developmental to control the movement of departing and arriving aircraft on the airport surface, including ramps and taxiways. Local control developed the skills to control arriving and departing aircraft on the active runways and in the immediate visual airspace of the terminal. Radar control taught techniques and procedures for the control of aircraft arriving in and departing from the terminal's extended

airspace for facilities equipped with radar. Data on the initial radar associate and initial radar qualification phases of training were collected for en route subjects. The en route radar associate phase qualified the developmental controller to initiate and accept radar hand-offs and point-outs, perform flight data entries, maintain flight progress strips, and communicate with aircraft and other facilities by interphone and radio *as directed by* the radar controller on a position. In contrast, the goal of the radar qualification phase of instruction was to qualify the developmental as the radar controller on two positions or sectors within the assigned area of specialization. The radar controller has overall responsibility for the safe, orderly, and expeditious movement of air traffic within the assigned sector of airspace. Performance assessments from additional radar training conducted by the FAA Academy were also extracted from the research data bases where available for subjects. FAA Academy radar training provided instruction in critical radar techniques and procedures in the safety of a simulated airspace.

An overall standardized composite score for each of 297 participants in this validation study was created from these time-to-complete, performance assessment measures, and FAA Academy radar training. This training performance (*TRNGPERF*) composite criterion represented the rate and quality of progress in training for an individual relative to peers assigned to the same type and level of facility that had completed the same curriculum. The mean *TRNGPERF* score was 0.44 (*SD* = .30), with a range of 0 to 1. A criterion score of 0 indicated consistently poorer (longer than average times to complete and lower assessments of quality). A score of 1 reflected consistently higher performance than peers (shorter than average times and higher assessments); an intermediate score of .50 indicated consistently average performance relative to peers assigned to the same type and level of facility.

Results

Correlations were computed between the current predictor *SCREEN*, alternative ATCS/PTS predictors, and the criterion. The correlation matrix was corrected for explicit and incidental restriction in range due to prior selection of the sample on the

current *SCREEN* predictor (see Ghiselli, Campbell, & Zedeck, 1981) and submitted for regression analysis. The corrected multiple correlation between the ATCS/PTS average final scores and *TRNGPERF* was $R = .25$ (uncorrected $R = .21$, p .05) compared to $R = .19$ (uncorrected $R = .11$, p .05) for the current *SCREEN* predictor. While modest, the validity coefficient of .25 for the ATCS/PTS indicated that a prediction about probable performance in field training for an individual could be made from knowledge of his or her scores on the computerized test battery. Moreover, the validity of the proposed 5-day test battery was at least equal to that of the existing 9-week training-as-screen. Subsequent analyses again suggested that the validities of the ATCS/PTS and ATCS Screen did not vary as a function of sex or minority group status (Weltin, *et al,* 1992).

Study 3:
Comparison of ATCS/PTS to job attribute requirements

A third analysis (Broach & Aul, 1993) of the ATCS/PTS was undertaken after it was validated in order to independently compare test constructs with job cognitive attribute requirements. During the data collection phase of the second study, FAA psychologists and technicians interviewed 52 of the incumbent FPL controllers from all types and levels of air traffic control facilities. Example facility types included Air Route Traff Control Centers (ARTCC), also known as En Route centers, Terminal Radar Approach Control (TRACON) terminals with high traffic densities, Level 3 radar terminals (L3R) with intermediate traffic densities, and Level 1 and 2 Nonradar (e.g., VFR Nonapproach) towers (L12NR) with lower traffic counts. The job analysts then completed a Position Analysis Questionnaire (PAQ; McCormick, Mecham, & Jeanerett, 1977) for each interview. Estimated requirements for worker attributes of an ability or aptitude nature were computed from the 52 sets of job ratings by PAQ Services, Incorporated, based on their data base on over 2,000 jobs in the U.S. economy. Preliminary data from this third analysis in the form of estimated percentiles for cognitive and general intelligence attributes are illustrated in Figure 4 for two selected air traffic control facility types and levels. The
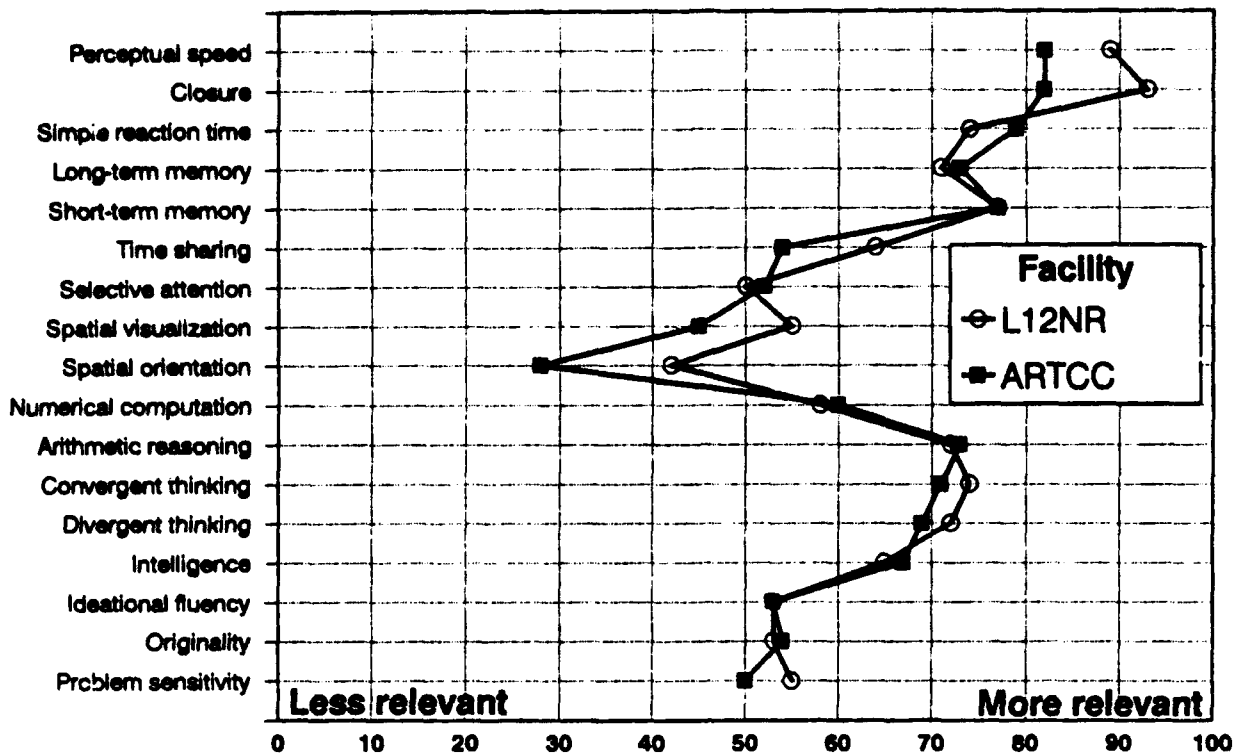
FIGURE 4. ATCS JOB ABILITIES PROFILE. *Ability attributes are listed along the vertical axis. The 0 - 100 scores along the horizontal axis indicate the estimated proportion of jobs in the PAQ Services, Inc. data base for which an attribute received the same or lower relevance scores than ATCS jobs. Ability requirements for the ATCS job in Level 1 & 2 nonradar/nonapproach (VFR) towers are contrasted with the profile for en route Air Route Traffic Control Centers.*

percentile estimates the proportion of the jobs in the PAQ data base for which an attribute received the same or lower relevance scores as the job being analyzed (Mecham & McCormick, 1969; Mecham, McCormick, & Jeanneret, 1977; McCormick, Jeannerett, & Mecham, 1972).

These analyses by Broach and Aul (1993) suggested that perceptual speed, closure, simple reaction time, and short-term memory were more relevant to the controller job than to many other jobs in the U.S. economy. Numerical computation, arithmetic reasoning, convergent and divergent thinking also appeared to be more relevant to performance in the ATCS occupation. But contrary to expectation, time sharing, selective attention, spatial visualization, and spatial orientation were not more relevant to air traffic cotnrol than to other U.S. occupations. In other words, there appears to be a substantial proportion of jobs in the

U.S. economy to which spatial and attention allocation abilities are more relevant than to the controller occupation. Finally, with the exception of spatial abilities as illustrated in Figure 4, the cognitive abilities requirements for controllers appeared to be reasonably homogenous across facility types and levels. The requirement for spatial visualization appeared to be more relevant to terminal facilities than to en route facilities.

Overall, tests that represented perceptual speed, closure, reaction time, memory, arithmetic reasoning, and some degree of spatial ability would be expected to predict performance in both en route and terminal environments. In order to evaluate the correspondence between test and job requirements, PAQ ratings of the proposed test battery were completed by a single, highly experienced PAQ consultant from Jeanerett and Associates. The resulting cognitive attribute re-
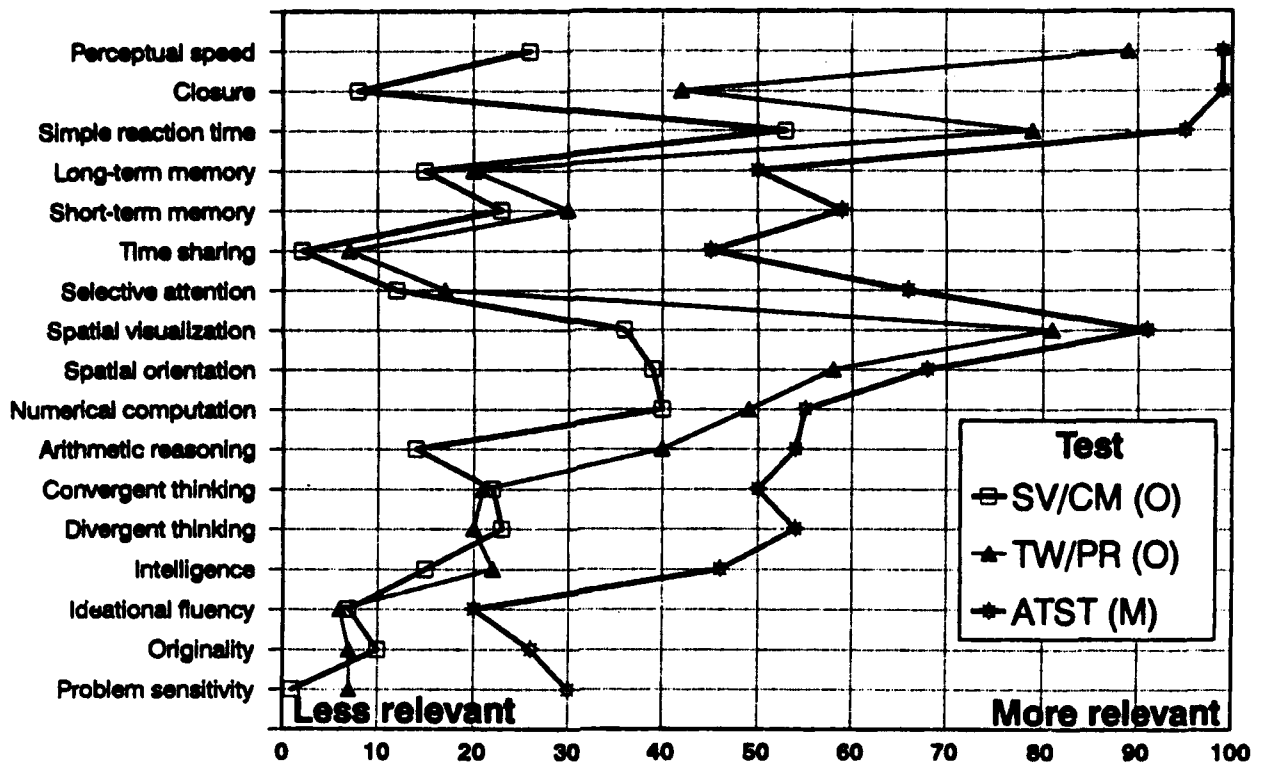
8

FIGURE 5. *ATCS/PTS ABILITIES PROFILE. Ability attributes are listed along the vertical axis. The 0 - 100 scores along the horizontal axis indicate the estimated proportion of jobs in the PAQ Services, Inc. data base for which an attribute received the same or lower relevance scores. Ability requirements for the SV/CM, TW/PR and ATST are illustrated.*

quirements for the test battery are illustrated in Figure 5 for this preliminary study of the correspondance between test and job. While no formal statistical analyses have been conducted as yet, there appeared to be some degree of similarity between the test and job profiles in kind, if not degree. For example, the requirement for perceptual speed and simple reaction time were similar between the ATCS job and the TW/PR and ATST tests. While the attribute percentile scores for the ATCS/PTS were generally lower, the shape of the profile across basic mental abilities such as memory and attention and higher-order skills such as numerical computation and divergent thinking was broadly similar to that of the job. Overall, these early data suggested at least some degree of correspondence between proposed test battery and job attribute requirements; further analyses, using multiple raters to evaluate the test battery, will provide a basis for a more definitive assessment.

DISCUSSION

Two formal validation studies on a total of 720 subjects demonstrated that the ATCS/PTS was a viable replacement for the ATCS Screen as the 2nd hurdle in the FAA's ATCS selection system. The first study demonstrated that the computer-administered test battery explained some of the variability in scores earned in the ATCS Screen, even after taking into account student aptitude. The second study found that ATCS/PTS was about as valid as the ATCS Screen in predicting relative performance in ATCS field technical training. The new test battery was objectively administered and scored, and the validity of the new test battery did not appear to vary as a function of sex and minority status. Finally, the ATCS/PTS achieved the major policy goal of reducing the cost of selection at the 2nd hurdle in the ATCS selection process from about $10,000 to about $2,000 per candidate. Therefore, the

9

FAA Academy ATCS Nonradar Screen was terminated in March 1992 and the ATCS/PTS became operational as the FAA's 2nd stage selection test in June 1992 on the basis of the results of the second. concurrent validation study. The ATCS selection system now consists of the 4-hour written ATCS aptitude test battery followed by, for those applicants earning a qualifying score, second-level screening on the ATCS/PTS. The final ATCS/PTS protocol provides 20 SV/CM, 20 TW/PR, and 20 ATST practice sessions over 2.5 days (Monday afternoon through Wednesday), followed by the final 4 SV/CM, 4 TW/PR, and 6 ATST "for grade" testing sessions on Thursday. Candidates are informed of the outcome of screening on Friday. Those that successfully complete the ATCS/PTS are then eligible for hiring by the FAA and subsequent enrollment in the FAA Academy ATCS training programs. In this new system, all selection is accomplished prior to the actual hiring and subsequent training of entry-level controllers.

The ATCS/PTS represents a major policy and research initiative for the FAA. As noted by Ackerman (1991), ATCS selection research represents a praxis of public policy, psychological theory, and psychometric practice. Continuing research is required to assess the longitudinal fairness of the new battery in order to satisfy legal and human resource policy requirements. An additional research requirement is to develop and validate an expanded test battery. Only cognitive abilities are assessed by the current version of the ATCS/PTS. But non-cognitive factors such as biographical data have been shown to be useful predictors of near-term criteria such as the ATCS Screen (Collins, Nye, & Manning, 1992) and criteria such as performance in radar-based training 1 to 2 years after entry into the occupation (Broach, 1992). Personality has similarly shown promise in several studies as a predictor of near-term performance (Schroeder, Broach, & Young, 1992; Nye & Collins, 1991). Development of a expanded test battery might enable the agency to implement a single-hurdle selection system, further reducing the financial costs of ATCS selection. A third important research requirement is the development of appropriate measures of ATCS job performance. What is validated in personnel selection research is the hypothesis that job performance, or important aspects of job performance, can be inferred from test scores (Guion, 1992). For example, given the nature of the criterion in the concurrent validation study, the only fully justified inference that currently can be drawn from ATCS/PTS scores is how rapidly a person might be expected to complete field ATCS training relative to other developmentals (slower or faster than average, overall). Inferences about probable technical job performance, such as efficiency in separating aircraft and orderliness of the flow of aircraft, will require development of different criterion measures. Similarly, inferences about attrition from the ATCS occupation from ATCS/PTS scores will have to await results of longitudinal evaluations of the Study 1 students and Study 2 developmentals as they progress through the field training program. Other important ATCS selection issues include differential assignment to facility types and levels based on test score profiles and assessment of selection system utility. Finally, as the controller occupation changes, the ATCS selection process must also change. The emerging Advanced Automation System may (or may not) have profound implications for ATCS selection (see Manning & Broach, 1992 for an early exploratory study). Systematic and continuous selection-oriented research is strongly recommended as an integral part of ATC systems design specifically and the national aviation human factors research plan generally.

REFERENCES

Ackerman, P. D. (1991). Discussion of selection of air traffic controllers: Complexity, requirements, and the public interest. In H. Wing & C. A. Manning (Eds.), *Selection of Air Traffic Controllers: Complexity, Requirements and Public Interest.* (DOT/FAA/AM-91/9). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Aerospace Sciences, Inc. (1991). *Air Traffic Control Specialist Pre-Training Screen Preliminary Validation: Final report.* (Final report delivered under contract DTFA01-90-Y-01034). Washington, DC: Federal Aviation Administration Office of the Deputy Administrator.

Aul, J. C. (1991). Employing air traffic controllers. In H. Wing & C. A. Manning (Eds.), *Selection of Air Traffic Controllers: Complexity, Requirements and Public Interest.* (DOT/FAA/AM-91/9). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Bittner, A. C. Jr. (1979). Statistical Tests for Differential Stability. *Proceedings of the Human Factors Society 23rd Annual Meeting* (pp. 541-544). Santa Monica, CA: Human Factors Society.

Broach, D. (1992). *Non-cognitive Predictors of Performance in Radar-based Air Traffic Control Training.* Paper presented at the 4th Annual Convention of the American Psychological Society, San Diego, CA.

Broach, D., & Aul, J. C. (1993). *Analysis of the air traffic control specialist (ATCS) occupation using the Position Analysis Questionnaire (PAQ).* Unpublished manuscript.

Collins, W. E., Nye, L. G., & Manning, C. A. (1992). Poststrike air traffic control trainees: Biodemographic predictors of success in selection and screening. *International Journal of Aviation Psychology, 2,* 213-223.

Della Rocco, P. S., Manning, C. A., & Wing, H. (1990). *Selection of controllers for automated systems: Applications from current research.* (DOT/FAA/AM-90/13). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Federal Aviation Administration. (1985). *National Air Traffic Training Tracking System.* (FAA order 3120.22A). Washington, DC: Federal Aviation Administration Office of Air Traffic Program Management.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences.* San Francisco, CA: Freeman.

Guion, R. M. (1992). Personnel assessment, selection, and placement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology: Volume 2* (2d Ed.) (pp. 327-397). Palo Alto, CA: Consulting Psychologists Press.

Human Technology, Inc. (1991). *Cognitive task analysis of en route air traffic control: Model extension and validation. Volume I: Report.* (Final report delivered under Office of Personnel Management contract OPM-87-9041). Washington, DC: Federal Aviation Administration Office of Training and Higher Education.

Lautenschlager, G. J., & Mendoza, J. L. (1986). A stepdown hierarchical multiple regression analysis for examining hypotheses about test bias in prediction. *Applied Psychological Measurement, 10,* 133-139.

Manning, C. A., & Broach, D. (1992). *Identifying ability requirements for operators of future automated air traffic control systems.* (DOT/FAA/AM-92/26). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Mecham, R. C. & McCormick, E. J. (1969). *The Rated Attribute Requirements of Job Elements in the Position Analysis Questionnaire.* (Report No. 1 under ONR Contract NONR-1100(28)). West Lafayette, IN: Purdue University.

McCormick, E. J., Jeanneret, P. R., & Mecham, R. C. (1972), A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). *Journal of Applied Psychology, 56,* 347-368.

McCormick, E. J., Mecham, R. C., & Jeanneret, P. R. (1977). *Technical Manual for the Position Analysis Questionnaire (System II).* Logan, UT: PAQ Services, Inc.

Nye, L. G., & Collins, W. E. (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: Federal Aviation Administration Office of Aviation Medicine.

Rodriquez, E., Narayan, K., & O'Donnell, R. D. (1986). *An Approach to Testing for Neurobehavioral Effects of Organic Solvents.* Los Angeles, CA: Atlantic Richfield Co.

Schroeder, D. J., Broach, D., & Young, W. C. (1992). *Contributions of Personality Measures to Predicting Success of Trainees in the Air Traffic Control Specialist Nonradar Screen Program.* Paper presented at the 1992 Annual Meeting of the Aerospace Medical Association, Miami, FL.

Sells, S. B., Dailey, J. T., & Pickrel, E. W. (Eds.) (1984). *Selection of air traffic controllers.* (DOT/FAA/AM-84/2). Washington, DC. Federal Aviation Administration Office of Aviation Medicine.

Shingledecker, C. A. (1984). *A task battery for applied human performance assessment research.* (AFAMRL-TR-84-071). Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory.

Siegel, A. I. (1978). Miniature job training and evaluation as a selection/classification device. *Human Factors, 20,* 189 - 200.

Siegel, A. I. (1983). The miniature job training and evaluation approach: Additional findings. *Personnel Psychology, 36,* 41 - 56.

U.S. Congress. (January 20, 1976). *House Committee on Government Operations Recommendations on Air Traffic Control Training.* Washington, DC: Author.

Weltin, M., Broach, D., Goldbach, K., & O'Donnell, R. (1992). *Concurrent criterion-related validation of the air traffic control specialist Pre-Training Screen.* (Final report delivered under contract DTFA01-89-Y-01019). Washington, DC: Federal Aviation Administration Office of the Deputy Administrator.

Wickens, C. D. (1984). *Engineering psychology.* Columbus, OH: Bobbs-Merrill Inc.