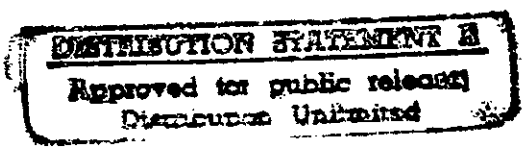
DOT/FAA/AM-97/15
Office of Aviation Medicine
Washington, D.C. 20591

Review of Air Traffic Controller Selection: An International Perspective

Dana Broach
Carol A. Manning
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

July 1997

Final Report



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

19970912 017

U.S. Department of Transportation
Federal Aviation Administration

DTIC QUALITY INSPECTED 3

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.

Technical Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-97/15		

4. Title and Subtitle	5. Report Date
Review of air traffic controller selection: An international perspective.	July 1997
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Broach, D., and Manning, C.A.	

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 Avenue, S. W. Washington, DC 20591	

15. Supplementary Notes

This research was conducted under task AAM-95-B-HRR-163

16. Abstract

This report provides a review of research on air traffic controller selection in the United States, Germany, the United Kingdom, and Sweden. The development and validation of the multiple hurdle selection system used by the US Federal Aviation Administration (FAA) from 1976 through 1992 is described first. The computer-administered test battery that supplanted the second-stage screening conducted at the FAA Academy is discussed next. Work by Eißfeldt for the German Air Navigation Services at the Aerospace Research Establishment (DLR) in Hamburg is reviewed. Job analysis, test battery development, and validation research for the controller occupation in the United Kingdom is presented next, followed by a description of the Swedish "MRU Project" on controller selection. The report closes with a discussion of issues surrounding controller job performance measurement. The advantages and disadvantages of historical criteria, such as training records, are reviewed. Alternative approaches to job performance measurement, such as simulations and operational data replay and analysis, are then described. The report closes with suggestions for future directions in controller selection research.

17. Key Words	18. Distribution Statement
Air Traffic Controllers, Performance, Personnel, Tests, Selection, Validation	Document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	27	

Form DOT F 1700.7 (8-72)

Reproduction of completed page authorized

REVIEW OF AIR TRAFFIC CONTROLLER SELECTION: AN INTERNATIONAL PERSPECTIVE

Choosing the wrong person for a job can have visibly disastrous results. Nowhere is this more apparent than in air traffic control, where the consequences of errors may be immediate and catastrophic. The method by which an organization selects the operators of intrinsically complex air traffic control systems is an important factor in achieving the goals of aircraft safety and efficient airport and airway management within increasingly constrained budgets. That method must take into account the nature of the air traffic control task, the range of human abilities relevant to performing the task, and the meaning and structure of performance. And, increasingly, that method must consider economics, public policy, and legal constraints. Air traffic control specialist (ATCS) selection, therefore, represents an intersection of public policy with psychological theory and research (Ackerman, 1991).

This chapter focuses on ways in which air traffic controllers are selected in organizations in the United States and other countries. The chapter is organized into 4 major sections. First, ATCS selection procedures in the United States are discussed. Next, current ATCS selection programs in Germany, the United Kingdom, and Sweden are briefly described. Third, the importance of developing criterion measures of controller performance is discussed. Finally, issues in ATCS selection research and research requirements are summarized.

ATCS SELECTION IN THE UNITED STATES

ATCS Selection 1976 - 1992

Development of a two-step selection process

Much of the research in ATCS selection has been conducted by the U. S. Federal Aviation Administration (FAA) since World War II (Hattig, 1991). Early research on the mental aptitudes required to succeed in training and on the job is described by Brokaw (1984). Written tests for the occupation were implemented in the United States as early as 1964, and included measures of arithmetic reasoning, spatial reasoning, and perceptual speed. Despite the increased validity of the civil service selection tests over a system based on prior experience alone, attrition at the Academy and in the field, as shown in Table 1, remained a substantive problem through the late 1960s and into the early 1970s (Cobb, Matthews, & Nelson, 1972; Henry, Kamrass, Orlansky, Rowan, String, & Reichenbach, 1975). As reported by Boone (1984), prior to 1971, attrition occurred during initial training at the FAA Academy, as well as in subsequent field training. However, pass/fail training at the FAA Academy was suspended in July 1971; as a consequence, field attrition rates increased dramatically, as documented in the 1975 Institute for Defense Analysis report authored by Henry, et al. In reaction to these increases in field attrition rates, the FAA re-implemented a centralized Initial Qualification Training course in 1976 to provide second-stage screening for en route and terminal controller candidates (Boone, 1984). Thus, since 1976, the ATCS selection process in the U. S. has consisted of 4 major steps: (a) a written aptitude test battery, (b) a personal interview, (c) a medical examination, including psychological evaluation, and (d) performance-based screening at the FAA Academy.

Research, as summarized by Collins, Boone, and VanDeventer (1984), continued through the late 1970s to improve the predictive validity and efficiency of the written test battery. For example, Buckley and Beebe (1970) developed the Controller Decision Evaluation (CODE) test, which consisted of a film of a computer simulation of air traffic movement across a radar scope. Subsequent studies demonstrated that the CODE test was a valid predictor of supervisory evaluations of field performance (Milne & Colmen, 1984) and attrition (Mies, Colmen, & Domenech, 1984). Translation of the CODE into paper-and-pencil format led to the Multiplex Controller Aptitude Test (MCAT; Dailey & Pickrel, 1984). The MCAT was designed to measure applicants' skills in applying a simplified set of ATC rules within a simulated air

Table 1

Historical ATCS attrition rates in the United States

Cohort	% Attrited at Academy	% Attrited in field training	% Retained in occupation
		Terminal Option	
1960-63	20.9	16.0	63.1
1968-70	19.3	10.1	70.7
1975		38.0	62.0
		En Route Option	
1960-63	32.0	22.8	45.2
1968-70	17.9	20.3	61.9
1975		43.0	57.0

Note: 1960-63 and 1968-70 cohort data from Cobb, et al; (1972) cohort from Henry, et al (1975)

traffic control environment. Studies confirmed that the MCAT was a valid predictor of performance in the FAA Academy (Rock, Dailey, Ozur, Boone, & Pickrel, 1981).

Based on this research record, a new ATCS selection battery was implemented in October 1981. This battery, administered by the U. S. Office of Personnel Management (OPM), consisted of the MCAT and the Abstract Reasoning Test (ABSR), which was retained from the previous civil service battery. The ABSR required the examinee to determine the relationships within sets of symbols or letters, and to identify either the next symbol or letter in a progression or the element missing from the set. Applicants also earned extra credit points based on their demonstrated job knowledge, as measured by a paper and pencil Occupational Knowledge Test (OKT; Lewis, 1978). The OKT was developed as an alternative to self-reports of aviation and air traffic control experience; it was found to be more predictive of performance in ATCS training than were self-reports (Dailey & Pickrel, 1984; Lewis, 1978). OKT and statutory veteran's preference points were added to transmuted MCAT and ABSR scores to yield a final overall civil service rating (Aul, 1991). Although the minimum qualifying score was much lower, a rating of 90 was usually required to be hired through the competitive civil service. Newly hired controllers reported to the FAA Academy for the second stage of screening.

Second-level screening of ATCS applicants at the FAA Academy began in 1976 and ended in 1992. Originally, the screening process included 2 programs, 1 for hires entering the en route option, and the other for hires entering the terminal option. In 1985, the 2 programs were consolidated into a single 9-week Nonradar Screen at the FAA Academy to reduce costs. The Nonradar Screen was designed to assess the aptitude of individuals having no prior knowledge of the occupation by teaching them a set of nonradar-based

air traffic control rules and principles, and then providing a series of laboratory simulation problems in which students demonstrated the application of those principles. Students completed the laboratory problems by performing the duties of an ATCS using nonradar procedures during standardized, timed scenarios encompassing the movement of aircraft through a specified airspace. During the problem, another student performed the roles of the aircraft pilots and other "controllers" participating in the scenarios. Instructors, former ATCSs trained to observe and rate student performance, graded the students' performance. Laboratory grades were comprised of 2 parts, the Technical Assessment (based on numbers and types of errors made) and the Instructor Assessment (based on the instructor's judgment of how well the student performed the problem, as compared with other students the instructor rated previously). A total of 13 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 of these post-1985 ATCS Screen performance measures was weighted 20% for classroom tests, 60% for laboratory scores, and 20% for the final examination. A minimum score of 70 out of 100 was required to pass. Candidates who did not successfully complete the ATCS Screen were removed from the controller occupation. Those who passed were assigned to a specific air traffic control facility for field training, and received a promotion. Trainee controllers, now termed "developmentals," required, on the average, 1.1 years (SD = 0.4) in non-radar, visual flight rules (VFR) towers, 2.2 (SD = 0.8) years in terminal facilities with radar, and 3.0 (SD = 0.6) years at en route centers (Manning, Della Rocco, & Bryant, 1989) to complete field training and attain the "full performance level" (FPL) of a certified controller.

Validity of the 1976-1992 two-step selection process

Several studies assessed the validity of the 1981 OPM written test battery and 1976-1992 ATCS Screen programs for predicting performance in field training. For example, VanDeventer (1981) found that the correlation between the composite score in the ATCS Screen and supervisors' rating of performance was .56 (adjusted for restriction in the range of predictor scores) for those in the en route option. At the time he conducted the study, no OPM test scores were available for analysis. Manning, Della Rocco, and Bryant (1989), found correlations of .46 (adjusted for restriction in range) between ATCS Screen score and both field instructor ratings and a measure of status in en route field training (based on whether a student had reached FPL status, was still in training, switched options, or failed). Similarly, Manning (1991a) found correlations of .30 and .44 (adjusted for restriction in the range of predictor scores) between field training status and OPM and ATCS Screen scores, respectively, for the 1986 cohort of ATCS Screen graduates. Multiple regression analyses found that the ATCS Screen score accounted for about the same percentage of the variance in field training status, as did the OPM score. A model containing the OPM rating alone predicted 12.5% of the variance in field training status, while adding the ATCS Screen score to the model predicted an additional 13.7% of the variance (adjusted for restriction in range). Broach and Manning (1994) found that ATCS Screen scores also had incremental validity over the written OPM test battery for predicting scores earned in both en route and terminal radar training taken 1 to 2 years after completing the ATCS Screen. These results suggested that, despite the apparent dissimilarities, a nonradar work sample predicted performance in radar-based air traffic control training. They called for additional research to elucidate the cognitive constructs underlying this empirical relationship between nonradar and radar air traffic control as part of the development of a new controller selection test battery.

ATCS selection 1992 - present

The multiple-hurdle selection process described above cost the FAA between $20 and 25 million annually to obtain approximately 1,400 trainee or "developmental" controllers to support rebuilding the ATCS workforce in the wake of the 1981 Professional Air Traffic Controller Organization (PATCO) strike. This selection process also imposed significant costs on applicants. Applicants selected to attend the

ATCS Screen had to leave their current jobs and, in some cases, families, for 9 weeks, with a 55 to 60% chance of remaining in the controller occupation at the end of the program. That risk may have discouraged potentially qualified women and racial minorities from pursuing an air traffic career (Aerospace Sciences, Inc. (ASI), 1991). The FAA undertook a major review of its ATCS selection program in 1990 to address these costs and other concerns. Three major ATCS selection policy goals were identified: (1) reduce the costs of ATCS selection; (2) maintain the validity of the ATCS selection system; and (3) reduce adverse impact on women and minorities. To achieve these goals, the FAA initiated the development and validation of a short-term, immediate replacement for the 9-week ATCS Screen, while at the same time, beginning longer-term research to support the Advanced Automation System. As only very preliminary conceptual studies are currently available for the longer term project, only the results of the short-term project are reported.

Development of a computer-administered test battery

Development of an interim computer-administered test battery to replace the ATCS Screen began in late 1990 by reviewing available information about the cognitive requirements of the ATCS job. Drawing on the available job analyses, such as a recent cognitive task analysis (Human Technology, Inc., 1991a), US researchers concluded that controllers primarily attend to multiple information sources, assess and integrate data, develop and prioritize plans of action, and implement those plans under time pressure while maintaining situational awareness. To assess the cognitive and sensory attributes required to perform these job functions, ASI developed a proposed test battery within the conceptual framework of 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 require performance of concurrent, multiple tasks by candidates to reflect the job demands placed on controllers.

Description of the computerized test battery

The 2 computerized information processing tests are (a) the Static Vector/Continuous Memory test (SV/CM; Figure 1) and (b) the Time Wall/Pattern Recognition test (TW/PR; Figure 2). Each of these tests consists of a pair of tasks, which are described in the figure legends. The SV component requires the subject to make judgments about conflicts, while the CM component exercises working memory.

The TW component is a time estimation task, while the PR task assesses perceptual speed. Subjects are presented with a fixed number of trials for a test within a nominal 5-minute SV/CM or TW/PR session; the actual length of the session is a function of subject response time. Performance feedback is provided at the end of each session on each test component. Measures from both the SV, CM, and PR components include the mean percent correct and mean reaction time for correct responses within the 5-minute sessions for each test pair; the TW measure is the absolute distance (in milliseconds) between the wall and target when stopped by the subject. The Air Traffic Scenario Test (ATST), the computer-administered work sample component of the proposed test battery, requires the subject to control aircraft within a simplified synthetic airspace, as described in the legend for Figure 3. Subjects direct aircraft to their destinations according to a small set of rules.

Aircraft are required to land at airports E and F at the lowest altitude and slowest speed, in the proper direction; while aircraft exiting gates A, B, C, and D must do so at the fastest speed and highest altitude. Aircraft at different altitudes are considered to be separated, while aircraft at the same altitude must be separated by at least 5 nautical miles, as represented by the separation icon. In addition, all aircraft must be separated from the airspace boundary by at least 5 nautical miles. Error counts are obtained and summed to create an overall error score. In addition, the system automatically computes the difference between the

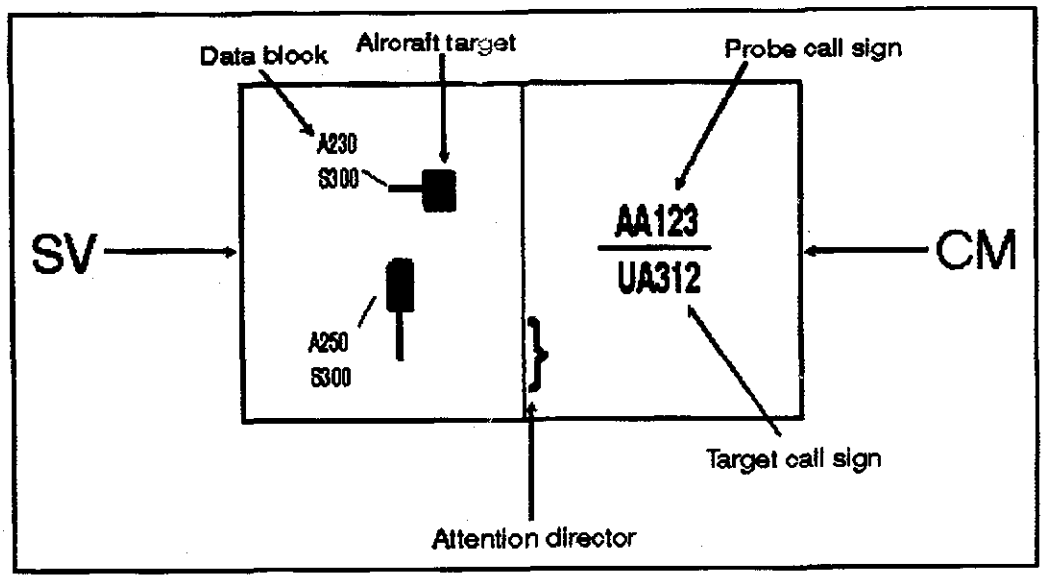


Figure 1: Static Vector (SV)/Continuous Memory (CM) screen.
SV test is shown on the left-hand side of the screen, CM test on the right. When the attention director was to the left, the subject's task was to decide if the aircraft targets would collide or not, based on the altitude ("A230") and speed ("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.

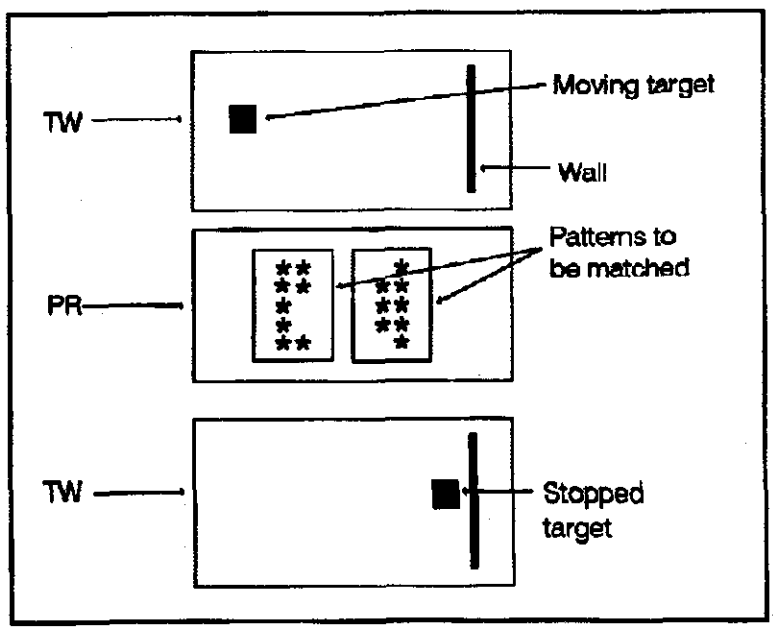


Figure 2: Time Wall (Tw)/Pattern Recognition (PR) screens.
First, the target appeared, moving from left to right at a steady speed toward the "wall" (*Top screen*). After an initial time interval, the target and wall were masked by a pair of patterns (*Middle screen*). The subject's task was to decide 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.

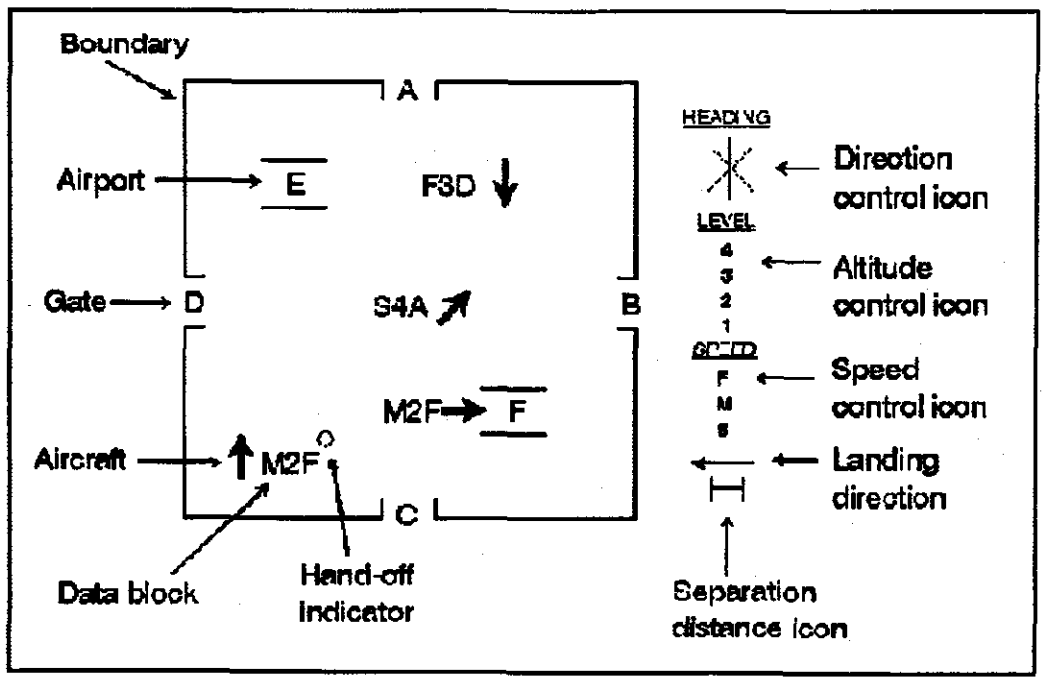


Figure 3: Air Traffic Scenario Test (ATST) Screen.
The boundary encloses a simplified airspace, with 4 outbound gates, *A*, *B*, *C*, and *D* and 2 airports, *E* and *F*. The aircraft and direction of flight are represented by the arrows adjacent to a data block. The alphanumeric data block indicates 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 subject 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.

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

Validation of the computerized test battery

Predictive, Criterion-related Validation. 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 OPM test battery. The sample in the first predictive, criterion-related validation study consisted of 423 newly-hired air traffic control students who entered the ATCS Screen in March and April 1991. The proposed test battery was administered to subjects the week prior to beginning the ATCS Screen. Instructions for 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 sessions. Performance feedback was provided to subjects after each session. On Friday, subjects received a final series of 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 Civil Aeromedical Institute (CAMI) research data bases after all subjects had completed the ATCS Screen. These data were matched with the proposed test scores for analysis; proposed test scores were not used in any way to make employment decisions about the subjects. The criterion for this predictive validation study was the final composite score earned in the ATCS Screen.

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. No corrections for restriction in range due to prior selection of the sample on aptitude were made for this analysis. First, the civil service rating at hire was entered into the regression equation predicting ATCS Screen score (R = .23, $F(1,357) = 19.23$, $p \leq .001$). Second, the proposed computer test scores were regressed on the criterion, using the SPSS-X (SPSS, Inc., 1988) forward stepwise variable selection method. The optimal linear combination of proposed test scores accounted for an additional 20% ($\Delta R^2 = .20$, $\Delta F(5,353) = 24.18$, $p \leq .001$) of the variability in final ATCS Screen scores beyond that already explained by student aptitude scores. There were no statistical differences in the prediction equation by sex and minority status (ASI, 1991).

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 an immediate replacement for the ATCS Screen (Weltin, Broach, Goldbach, & O'Donnell, 1992). The sample for this second validation study was composed of 297 trainee ("developmental") and fully trained and certified "full performance level" (FPL) controllers. 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 ATCS 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 to be evaluated. 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 were used in the first study.

Data describing progress in training were combined to create a composite criterion for validating the ATCS/PTS. The source data included the number of days spent in particular phases of field training and hours of formal, documented on-the-job training (OJT), as reported by field ATC facilities in accordance with national policy (FAA, 1985), and subjective ratings of developmental performance in that phase of training by instructors or supervisors. Scores earned in radar training at the FAA Academy were available for many subjects as well. 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 measures, as described in Weltin, Broach, Goldbach, and O'Donnell (1992). This training performance composite criterion represented the rate and quality of progress in training for an individual, relative to peers who had completed the same curriculum and were assigned to the same type and level of facility. The mean training performance criterion 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.

Correlations were computed between the current predictor (the FAA Academy ATCS Screen final score), alternative predictors (ATCS/PTS scores), and the criterion (a composite of standardized scores for field training performance). The correlation matrix was corrected for explicit and incidental restriction in range due to prior selection of the sample on the current predictor (see Ghiselli, Campbell, & Zedeck,

1981) and submitted for regression analysis. The corrected multiple correlation between the ATCS/PTS average final scores and the training performance criterion was $R = .25$ (uncorrected $R = .21$, $p \leq .05$) compared to $R = .19$ (uncorrected $R = .11$, $p \leq .05$) for the current 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 ATCS 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, Broach, Goldbach, & O'Donnell, 1992).

Discussion of the FAA computerized test battery

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 second hurdle in the FAA's ATCS selection system. The predictive 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 concurrent 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 per candidate selection cost at the second hurdle in the ATCS selection process from about $10,000 to about $2,000.

The FAA Academy ATCS Nonradar Screen was terminated in March 1992 and the ATCS/PTS became operational in June 1992 on the basis of the results of the 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.

INTERNATIONAL ATCS SELECTION

Controller selection is an equally important human factors issue for air traffic control (ATC) systems outside the United States, particularly with the increasing internationalization of air travel and ATC systems. Controllers in Germany, for example, must control flights that may cross multiple national boundaries, requiring coordination with controllers in Sweden, Italy, Switzerland, and France. Demographic trends, with aging employees in some cases, increasing traffic loads, and technological innovations in ATC systems, requiring expansion of controller staffs, are creating unique demands on controller selection throughout the world. In this section, controller selection systems and supporting research in Germany, the United Kingdom, and other countries are briefly described, based on available reports. International ATCS selection research has been reviewed by Hilton and Sells (1984), and more recently by Hattig (1991). Hilton and Sells concluded that standardized ATCS qualification and licensing might be necessary, in view of the continued expansion and integration of ATC systems and increased job complexity. Hattig focused specifically on military controller selection in several European countries. This summary focuses on civilian controller selection procedures. We hope that additional detailed information about these, and other selection systems, will become more readily available and more widely shared within the research community.

ATCS SELECTION IN GERMANY

Research supporting the selection of controllers in Germany is conducted by the Department of Aviation and Space Psychology in the German Aerospace Research Establishment ("DLR") for the Federal Administration of Air Navigation Services (BFS). The BFS is the German counterpart to the US FAA; the DLR is, in many respects, parallels the FAA Technical Center and the FAA Civil Aeromedical Institute in the area of ATC research, engineering, and development. The applied research goals of the research program in the Department of Aviation and Space Psychology are to (a) develop selection procedures to predict performance in ATCS training, and (b) ensure that the controllers are able to cope with the high ATC job demands until retirement (Eißfeldt, 1991).

Description of German selection process

The DLR has developed and validated a 4 step ATC selection procedure that requires about 4 days to administer:

1. A "Pre-Selection" phase consisting of a battery of 8 paper-and-pencil tests;
2. Part I of the "Main Selection" phase, consisting of 11 additional group-administered, paper-and-pencil tests plus a test of vigilance;
3. Part II of the "Main Selection" phase, consisting of apparatus tests plus an oral English language examination;
4. Part III of the "Main Selection" phase, consisting of an interview with a board comprised of a senior controller, 2 other experienced controllers, and 2 DLR aviation psychologists as advisors.

Seven cognitive and 9 personality traits are assessed in course of the DLR selection procedure (Table 2). The personality traits were measured with the Temperament Structure Scales (TSS; Goeters, Timmerman, & Maschke, 1993). The instruments or procedures used to assess the traits in the "performance domain" are not identified or described in the available English-language reports. Nor is it clear from these reports which traits are assessed in what stage of the selection process, and by what instruments. Eißfeldt indicated that approximately 40-45% of German applicants proceed to the main phase of the selection process, with about 10% of the total applicant group successfully completing both phases of the DLR selection procedure. For example, just 644 of the 8,646 applicants completed both phases, for a net selection rate of 7.4% during the period 1982-1988. Similarly, 11,280 persons out of 238,946 applicants in the United States successfully completed both the first

Table 2

Traits assessed in the German ATCS selection system

Performance Domain	Personality Domain
Basic knowledge	*Achievement-oriented traits*
English	Motivation
Technical Comprehension	Rigidity
Mathematico-Logical Thinking	Mobility
	Vitality
Operational Attitudes	*Interpersonal Behavior*
Memory	Extroversion
Perception & Attention	Dominance
Spatial Orientation (Auditory/Visual)	Aggressiveness
Multiple Task Capacity	Empathy
	Stress Resistance
	Emotional Stability

stage written aptitude examination and second stage work sample at the FAA Academy, for a selection rate of about 5% during the same period.

Validity of the German selection process

As in the US, the German ATCS selection test battery has been validated against training outcomes, rather than measures of core ATCS technical performance. For example, Eißfeldt reported an attrition rate of 18% (36 of 475 trainees) in subsequent ATCS training, which is about the same as the US loss rate in terminal training, but less than the attrition rate in its en route centers. The validity of the German battery was also assessed in a sample of 201 controllers by examining the relationships between predictor test scores and the following training success criteria:

- Written examination at 6 months on aspects of law and civil service;
- Performance tests at 24 months on ATC problems in a radar simulator;
- Performance tests at 34 months in 3 different working ATC positions;
- Final examination average score in all theoretical and practical aspects of ATC; and
- Overall pass or fail in training.

The 20 unweighted test results from the pre- and main phases of the DLR selection procedure were used to predict training outcomes for each criterion in a series of discriminant analyses. Validities (*R*s) of the test battery against the examination and test criteria ranged from .51 to .61 (all significant), resulting in 67 to 78% correct classifications with respect to criteria such as 2 written examinations (pass/fail; $R = .55$ and .51), a radar simulation (pass/fail; $R = .61$), and final grades in training ($R = .61$). Sample sizes ranged from 162 to 196 entry-level controllers. In comparison, Broach and Manning (1994) reported a multiple correlation coefficient (*R*) of .50 between scores on the first and second stages of the US ATCS selection process and performance in en route radar training. However, the regression of the German selection tests scores on overall pass/fail status in training was not significant. In comparison, Manning (1991a) obtained an R of .27 ($N = 402$) between FAA selection test scores and training outcome without adjustments for restriction in range. Both German and FAA en route training required an average of about 3 years to complete.

ATCS SELECTION IN THE UNITED KINGDOM

The ATCS selection process in the United Kingdom was based on existing civil service qualifications up until the mid-1980s. With small numbers of trainees required each year (as few as 50 per year) and a 3 year apprenticeship, that selection procedure met the needs of the UK ATC system, despite an attrition rate as high as 49% in 1984 to 85 (Browne, 1993). However, a review was undertaken in 1983, at the behest of the UK Civil Aviation Authority in anticipation of increased manpower requirements in the 1990s. The project consisted of a job analysis and concurrent test validation study.

Job analysis

The ATCS job analysis was conducted in late 1982 by Saville and Holdsworth, Ltd. (SHL). The occupational psychologists from SHL used a variety of techniques, including the Position Analysis Questionnaire (PAQ), Critical Incidents Technique (CIT), and Repertory Grid Interviews (RGI) to elicit job information in a structured manner. The job analysts also reviewed relevant documents, observed controllers at work, and conducted interviews with key personnel (Nyfield, Kandola, & Saville, 1983). Their analysis resulted in the "tentative model" presented in Table 3. The core of controller skill appears, in the SHL analysis, to involve rapid processing of information from multiple channels in order to develop and maintain a "real-time" representation of events in the airspace. Controllers apply this skill, or set of skills, in a time-pressured, repetitive or cyclic work context in the presence of distractions. Application of these core skills, in this context, appears to require a self-confident, conscientious, and cooperative temperament.

Table 3

United Kingdom model of the ATCS job

Core skills	Contextual factors	Temperamental factors
Ability to absorb information simultaneously from multiple sources	Speed of decisions	Readiness to work within a system
Ability to absorb new information while making decisions	Sporadic time pressure	Preference for working to set standards
Ability to project forward on the basis of current information	Sudden high-level demands on the individual	Cooperativeness
Ability to constantly adjust the whole picture	Distractions	Convergent thinking
	Fluctuations between routine and non-routine	Decisiveness and confidence
	Checking/updating information	Conscientiousness
	Short-cycle repetitive work	Structured thinking
		Self-control

Test battery development

Cognitive ability tests. Six ability tests were developed by SHL on the basis of the job analysis to assess characteristics associated with controller performance. In the 10-minute Basic Checking test, the examinee was required to find the number or letter string from among 5 alternative strings on the right-hand page that exactly matched the probe string on the left-hand page. The Basic Checking test closely resembles the Number Comparison Test (P-2) of the *Kit of Factor-Referenced Cognitive Tests* (Ekstrom, French, & Harman, 1976), and appears also to be a measure of perceptual speed. The 10-minute Audio Checking test closely resembles the Basic Checking test, except that the stimulus string is presented orally. This unique test appears to assess both short-term memory and perceptual speed, and requires processing using both auditory and visual resources. In the 15-minute Visual Estimation, a series of 5 lines, angles, or figures are presented to the examinee in each item. The examinee's task is to identify the 2 lines, angles, or figures that are identical. This test is reminiscent of the Identical Pictures Test (P-3) of the Ekstrom, et al. set of tests, and perhaps offers a nonverbal assessment of perceptual speed. However, Nyfield, Kandola, and Saville (1983) describe the Visual Estimation test as a measure of spatial aptitude that is relatively independent of general intellectual capability. The Spatial Reasoning test (20 minutes) presents a pattern, which, when folded, creates a cube. As in the Surface Development Test (VZ-3) of the Ekstrom, et al. tests, the examinee must try to imagine, or visualize, how the object would look from a variety of perspectives when folded. This test appears to be a relatively pure measure of spatial visualization (e.g., the ability to manipulate visual images in 3 dimensions mentally (Mecham, McCormick, & Jeanneret, 1977)). The Diagramming test (20 minutes) is described as measuring "logical analysis through the ability to follow complex instructions" (Nyfield, et al., 1983, p. 7). The stimulus consists of 1 or more boxes arranged in a column on the left, paired with an equal number of

circles in a column on the right. There is a geometric figure, such as a half-shaded diamond, in the box, and a symbolic operator inside the circle. The symbolic operators are defined for the examinee in a separate list. The figures in the boxes are changed in a specified way by the symbolic operators in the circles. The examinee's task is to choose, from among the 5 alternatives, the column of boxes resulting from carrying out the operations described by the stimulus. This test may represent a measure of non-verbal general reasoning ability; however, there is no clear analogue in any of the factor-referenced cognitive tests described by Ekstrom, French, and Harman (1976). The 15-minute Diagrammatic Reasoning test resembles the Abstract Reasoning component of the FAA written ATCS aptitude test battery, in which the examinee must determine the next figure in a series of figures in a logical sequence. Such tests may also assess a non-verbal general reasoning ability.

Personality test. Given the salience of temperamental factors in the ATCS job analysis, SHL also included a personality test in the test battery. The Occupational Personality Questionnaire (OPQ; Saville & Holdsworth, 1990), then under development, was designed to assess 32 personality factors, or traits, in 3 domains (Table 4). No estimates of internal consistency, test-retest reliability, or evidence of convergent and discriminant validity for the OPQ were provided in the Nyfield, Kandola, and Saville (1983) report. However, subsequent reports assert that the psychometric characteristics of the OPQ are at least as good as those of other widely used personality measures, such as the 16PF (Robertson & Kinder, 1993; Saville & Wilson, 1991; Swinburne, 1985).

Validation of the test battery

The sample for the 1983 validation study consisted of 154 incumbent controllers with between 4 and 10 years of service. The mean age of the sample was 30.8 years, with a range of 22 to 44 years; 70% of the sample was between the ages 27 and 32. The majority (76%) of the mostly male (96%) sample had become controllers after a period of service as an ATC assistant, with just 24% having entered controller ranks directly. The validation criteria in this study consisted of a set of 22 rating scales based on the model of controller job performance presented in Table 3. Each rater was asked to think of the "least good" and "best" controller as anchors for the rating scales, and rate each controller on an 11 point scale. Factor analysis of the 22 job performance rating scales yielded a 3 factor solution. Factor 1, interpreted by Nyfield, Kandola, and Saville (1983), as a general job performance factor, appears to represent core technical performance. The second factor might be interpreted as representing teamwork, and appears to be linked to the temperament aspect of the model of the controller job derived in the course of the job analysis. The third factor reflects situational awareness. Factor scores were computed for each subject, and used as criteria in a series of multiple regression analyses to explore the validity of the proposed paper-and-pencil test battery.

Full and complete predictor and criterion data on 112 subjects were available for the validity analyses. As a first step in the validity analysis, bivariate correlations between predictors and job performance scales were examined. In general, subjects rated more highly overall also had higher scores on the Basic Checking (perceptual speed; r = .17) and Diagramming (non-verbal logical reasoning; r = .22) tests. On one hand, they also appeared not to be looking for aesthetic qualities in their work (r = -.19 with OPQ Artistic), nor to be status conscious (r = -.21 with OPQ Leading). On the other, the subjects appeared to be modest (r = .21 with OPQ Modest) and accepting of facts and assumptions (r = -.25 with OPQ Critical), . The second step was to regress test scores on the factor scores. Scores on the Diagramming test, OPQ Critical, and OPQ Competitive scales accounted for 17% (R = .42) of the variance in the general performance or core technical performance factor (Factor 1). Nyfield, et al. (1983) concluded that the 2 tests drawn from the spatial domain of abilities (Visual Estimation and Spatial Reasoning) were not related to the performance of incumbent controllers, and recommended that they not be included in the operational test battery. They recommended inclusion of the Basic Checking and Diagramming tests in the operational test.

Table 4

Occupational Personality Questionnaire (OPQ) domains and factors (United Kingdom)

Relating Domain	Thinking Domain	Feeling Domain
Leading	Forward thinking	Relaxed
Competitive	Conservative	Optimistic
Modest	Practical	Emotionally controlled
Socially confident	Detail conscious	Self-aware
Caring	Data rational	Achieving
Independent	Critical	Worried
Persuasive	Conscientious	Phlegmatic
Democratic	Conceptual	Self-esteem
Effusive	Innovative	Active
Tolerant	Tolerant of ambiguity	
Gregarious	Artistic	
	Decisive	

Discussion of the United Kingdom selection battery

The test battery was implemented in the mid-1980s, and is the subject of continuing research and evaluation. Results from the most significant tests are combined with selected OPQ scales to compute a sten score, which predicts the probability of success, if selected for controller training (Browne, 1993). Currently, applicants with sten scores of at least 6 (41% of applicants) are eligible for the interview phase of the United Kingdom Civil Aviation Authority selection process. However, a large applicant pool generated in 1991 (some 3,500 in the first 3 months of 1991) has allowed top-down hiring, with sten scores of 8 or above (25% of applicants) required for referral to the interview (Browne). A computer-based test battery is currently under development, with the intention of assessing additional ability constructs not easily captured by paper-and-pencil tests.

Overall, the magnitude of the correlations reported by Nyfield, Kandola, and Saville (1983) are comparable to those reported in the United States and Germany for their selection systems. Perceptual speed and nonverbal logical reasoning appear to be important predictors of technical job performance, as suggested by the Broach and Aul (1993) job analysis. The results for the spatial measures in the United Kingdom provide empirical support for the counter-intuitive job analysis findings by Broach and Aul in the U. S. and Mogford and Tansley (1991) in Canada, that spatial abilities *per se* may have less relevance to the performance of ATC tasks than previously supposed.

ATCS SELECTION IN SWEDEN

The Swedish Civil Aviation Administration (Luftfartsverket; LFV) was created in the late 1970s by the integration of the Swedish military and civilian ATC systems. One goal of that merger was the creation of an ATCS selection and training program with a maximum failure rate of just 20%. However, as reported by Haglund (1993), that objective has not yet been achieved, as the pass rates range between 63% and 89% for the present LFV ATCS selection and training process. The selection component of LFV process is based on a series of tests and interviews, where choice of the tests appears to have been made on the basis of a general analysis of the controller job

(Haglund, Andersson, Backman, Brehmer, & Sundin, 1993). Initially developed in 1978, the aptitude tests used by the LFV were grouped into 4 general factors (Table 5): (1) Flexibility and ability to find new solutions; (2) logical reasoning ability; (3) spatial ability; and (4) attention to detail, carefulness, and short-term memory. The Defence and Object Relations Test (DORT; Svensson & Trygg, 1991) projective personality test is administered to candidates passing the written test battery. The LFV Air Navigation Services Department recently undertook a major project to lay the foundation for a new controller selection system by (a) examining the validity of its current tests, and (b) conducting a more specific job analysis based on the Swedish ATC system, as reported by Haglund, Andersson, Backman, Brehmer, and Sundin (1993).

Validity of existing tests

Complete test and training data were available for 134 students admitted to ATCS training in 1990 and 1991. Regression analyses were performed to assess the relative validities of the test battery components described in Table 4, using training outcome (pass or fail) as the criterion. Haglund, Andersson, Backman, Brehmer, and Sundin (1993) reported an overall multiple correlation of .413, but noted that the adjusted squared multiple correlation was just .032 (*ns*). Reduction of the predictor set through stepwise regression resulted in an R of .334 (adjusted R^2 = .091, $F(3,130) = 5.44$, $p \leq .01$). The reduced predictor set consisted of the Skeppsdestination ("Ship's Destination") and Korrektur ABAR ("Proof-reading ABAR") tests, and interview by LFV personnel. The interview was negatively weighted; that is, applicants receiving a higher rating in the interview were less likely to succeed in training. The interview was described as an assessment of a candidate's ability to cope with stress, to cooperate, to take the initiative, and motivation to become an air traffic controller. The Skeppsdestination test was described as assessing flexibility - the ability to find new solutions or to improvise and make decisions in unexpected situations. The Korrektur ABAR test was described as measuring attention to detail, carefulness, and short-term memory.

Job analysis

Structured interviews were the primary data collection tool in the Swedish study. A total of 127 incumbent controllers participated in group interviews at 11 ATC and 2 training sites. Participants were asked about the behaviors used by skilled controllers to cope effectively with stressful situations or events. Effective behaviors were tabulated by stressful event, resulting in an event-by-behavior table, with frequency of occurrence of an effective coping behavior for a stressful event ($f_{event, behavior}$) as the cell entries. This tabulation of behaviors and stressful events was used to develop questionnaires tailored to each operational environment (tower, approach control, and en route). The questionnaires were then distributed to a representative sample of 158 controllers and instructors. Participants were asked to rate, on a 7-point scale, (a) the importance of each coping behavior, and (b) how often the related stressful events or situations occurred in daily work.

The stressful situations, or events, were grouped into five broad categories by the Swedish researchers: (1) traffic processing; (2) coordination; (3) disturbances and irregularities; (4) fluctuating workload; and (5) personalities and social skills. Similarly, the effective coping behaviors were also sorted into 5 categories: (1) decision-making; (2) self-confidence; (3) information gathering and processing; (4) social relations; and (5) communications. However, only a narrative of the results is presented, rather than detailed statistical tables describing the frequency of occurrence of events or the rated importance of coping behaviors, either overall or by ATC environment. The authors do indicate that "information gathering and processing" behaviors were most effective in relation to coordination and traffic processing events in en route control centers. Decision-making and communications behaviors appeared to be most important to coordination and traffic processing events or situations in the approach control environment. Finally, decision-making, information gathering and processing, and self-confidence seemed to be more important to coping with stressful traffic processing situations in the tower environment.

Table 5

Swedish Civil Aviation Authority selection test battery

General Factor	Description	Tests	
		Swedish Name	English Name
Flexibility and ability to find new solutions	Ability to improvise and make decisions in unexpected situations	**Skeppsdestination**	Ship's Destination
		Instrutionsprov II	Instruction Test II
		Kravatt	Neck Tie
Logical ability	Logical ability		Raven's Progressive Matrices
			Raven's Number Series
Spatial ability	Ability to construct a 3-dimensional picture of the airspace from 2-dimensional information	**Klossar**	Blocks
		Platmodeller	Metal Sheet Models
		WIT Puzzles	WIT Puzzles
Attention to detail, carefulness, short-term memory	Attention to detail, carefulness, short-term memory	**Korrektur ABAR**	Proof-reading ABAR
		Sifferkorrektur	Number proof-reading
		Namnminne	Name memory
		Sifferminne	Number memory
		Figuridentifikation	Figure identification

Future research and development in Sweden

Based on the results of the validation of the current LFV selection tests for controllers, research on alternative tests continues. For example, the LFV has inquired about collaborative research with the FAA using the new computerized ATCS/PTS test battery (J. Aul, personal communication). Haglund (1993) noted the importance of linking dynamic assessments of relevant abilities in the new selection process to the important groups of job behaviors, as has the FAA, Germany, and the United Kingdom. A critical issue, however, in the development and validation of a new generation of tests for all countries, is measurement of controller job performance.

ATCS JOB PERFORMANCE MEASUREMENT

Measuring the ATCS performance is critical, for as Guion (1992) noted, 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. Controller selection research has relied upon training success as *the* criterion for validation of tests, rather than job performance. However, a fundamental question is *what* is meant by "controller job performance?" Job *performance* may refer to either (a) the temporal sequence and experience of interlocked and covariant individual behaviors

necessary to the achievement of organizational goals and objectives, or (b) the results or consequences of those individual behaviors relative to organizational goals (Binning & Barrett, 1989; Weick, 1979). Job performance *measurement*, however, is the scaling of behaviors, products, and services in terms of success or failure relative to organizational standards, goals, and objectives. Therefore, while the tools of industrial/organizational psychology and human factors are invaluable to performance measurement, the definition of the referent standards, goals, and objectives against which behavior is to be scaled remains, and must be, the responsibility of the air traffic system managers. System managers must carefully define what aspects of controller behavior are important and to be explicitly predicted from selection test scores. In addition to the requirement to develop criteria that reflect management's priorities, the measures must also be practical, reliable, and valid (Siegel & Lane, 1982). A single criterion is probably not appropriate, as performance is most likely to be a multidimensional construct (Campbell, McHenry, & Wise, 1990).

Another question to be considered is at what point in a controller's career should criterion data be collected? For example, military pilot selection tests have often been validated against flight training outcomes about 1 year after entry into training (Hilton & Dolgin, 1991). Similarly, as the job of a developmental controller for 1 to 3 years is to learn the job, training status (successful completion or not) has often been used as a criterion for validation of written ATCS aptitude tests, the FAA Academy Screen, and the computerized ATCS Pre-Training Screen. However, one might argue that training performance is not equivalent to (or may not even be highly correlated with) job performance, depending on the type of training provided. Static measures of training performance, such as paper-and-pencil knowledge tests, might not correlate highly with performance on a highly dynamic job, such as air traffic control. But, if the training measures were obtained in a sufficiently dynamic environment and were sufficiently sensitive, it might be appropriate to use training measures as surrogates for job performance. Yet, 2 to 3 years, on the average, is required to complete ATC training in the United States; as a result, initial performance may have little relationship to performance in later stages of training. Criterion measures obtained at different intervals over an extended time are often poorly correlated (Siegel & Lane, 1982). Moreover, given several years of training and job experience intervening between predictors and actual FPL job behavior, it might be expected that the link between predictors and FPL job performance measures would also weaken over time. Because of such time delays, it may not be appropriate to use job performance as an ATCS selection test validation criterion, although that relationship would continue to be of interest to the researcher.

Finally, another important factor that should be taken into account when choosing or developing criterion measures is whether they measure typical or maximum performance (Sackett, Zedeck, & Fogli, 1988). Does the performance of a controller when he/she is expending a maximal effort or performance on a typical day better represent the type of performance that we are trying to predict with a selection test? Sackett et al. proposed that while aptitude contributes to both types of performance measures, motivation also contributes to the measurement of maximum performance. In fact, the two types of measures were not highly correlated in Sackett et al.'s (1988) study.

Criterion measures used historically in air traffic control

A number of tests and biographical factors have been consistently predictive of success in US air traffic control over the years. The nature of the criteria used in these validation studies must be kept clearly in mind when interpreting these results. For example, Brokaw (1952) collected official evaluations, supervisors' unofficial ratings, biographical/career data, and checklists of effective and ineffective behaviors on a sample of 210 incumbent Airway Operations Specialists (AOS, the predecessor occupation to today's controllers). The measures were essentially supervisors' global assessments of the quality of controller performance. These measures, taken after training program completion, may have represented the supervisors' impression of typical performance.

Training measures

In contrast, measures derived from the FAA Academy *ATCS* Screen were used as validation criteria during the late 1970s and early 1980s. These measures were considered appropriate because, while the ATCS Screen was a selection procedure, it was at the same time a type of work sample test that assessed performance on a task that resembled the job of controllers in many important ways. While the ATCS Screen was in existence, all test (and item) scores and ratings earned on laboratory problems were recorded, as well as a count of the numbers of different types of errors committed on each problem. The measures derived from the tests and laboratory problems reasonably described student performance, had desirable distributional properties, and were also readily available. In spite of their apparent advantages, several problems were associated with the use of these scores as test validation criteria. First, because the Academy program was a selection procedure, the criterion measures were obtained at the beginning of a student's career, and thus did not measure how well students performed on the job. Furthermore, not all candidates had learned to perform the activities required during the problems at the time of testing; thus, performance was measured somewhere along the learning curve rather than at asymptote. Second, the scores obtained for the laboratory problems were based on 2 types of subjective judgments. One was an instructor's count of the types of errors committed and the other was a subjective rating of student potential. Third, the laboratory problems used in the Screen were based on nonradar procedures infrequently used in today's system. The argument that performance in nonradar control is predictive of performance in radar control was irrelevant to the use of those measures as criteria purported to represent *job* performance.

During the *1980s* and *1990s*, several measures of performance in simulation and on-the-job training were used as criteria. The first was a set of measures obtained from the Academy's Radar Training course, which were comparable to those measures obtained for the Academy screen program in that they represented performance measured by both paper-and-pencil tests and in laboratory simulation problems. For example, Broach and Manning (1994) used Academy radar training scores as criteria against which the incremental validity of the nonradar screen over that of the *OPM* test was determined. Essentially, these radar scores represented mastery of a relatively circumscribed set of radar knowledges and skills, and were collected at an intermediate stage of training from trainees rather than FPL controllers. Radar training scores were more acceptable criteria because they came from a program that taught the use of radar procedures utilized in today's operations. However, as in the nonradar Screen program, the scores for radar laboratory problems were based on instructor subjective judgments about errors and potential to succeed in the ATCS occupation. Other studies examined the validity of ATCS selection procedures relative to performance in field training (Manning, Della Rocco, & Bryant, 1989; Manning, 1992; Manning & Aul, 1992; Weltin, Broach, Goldbach, & O'Donnell, 1992). General information about training performance was collected for every phase of training for each controller trainee who entered field training. Several types of information on performance were obtained: the start and completion dates, the number of hours used to complete on-the-job training (OJT), and the grade (Pass, Fail, or Withdraw). A rating of controller potential, measured on a 1 to 6 scale, was made by an instructor or supervisor who most frequently observed the student during that phase. This information can be aggregated to derive measures of training performance, such as the amount of time (in years) required to reach journeyman or full performance level (FPL) status, instructor ratings averaged across combinations of training phases, time (in days) to complete OJT in certain training phases, and total number of OJT hours required to complete those phases. These measures could be interpreted as representing the rate and quality of progress through a resource-intensive apprenticeship (Weltin, Broach, Goldbach, & O'Donnell, 1992) through all stages of training. However, as with the Screen and radar training measures, these field training measures do not represent FPL job performance.

Moreover, the measures of field training performance had a variety of problems that limited their utility as criteria. The most notable problem was that a number of outside factors (besides aptitude and

technical proficiency) may have affected the accuracy of their measurement. For example, time to reach FPL status may have been affected by delays in training caused by the need to use the controller in an operational position, the number of other students undergoing OJT on the same airspace, the amount and complexity of airspace in the student's area of responsibility, the order in which different portions of the airspace were learned, and/or the availability of the training laboratory. The number of OJT hours used may have been affected by the number of OJT instructors who provided training to a student, and insufficient exposure to different types of air traffic during training. The subjective rating of student potential could be affected by a number of rating biases familiar to psychologists (e.g., leniency, central tendency, severity, halo effect, contrast and similarity errors; Siegel & Lane, 1982).

Personnel records

Career progression of air traffic control trainees is another type of criterion that has often been used by the FAA (VanDeventer, 1981; Manning, Della Rocco, & Bryant, 1989; Della Rocco, Manning, & Wing, 1991). Information to construct this type of measure has been obtained from the FAA's Consolidated Personnel Management Information System (CPMIS). CPMIS contains dates of entry, separation, and movement between facilities, changes in job series and grades. Typically, this information has been used to describe status in training at the first facility. Other types of measures that could be obtained from the CPMIS system are current government service (GS) pay grade, maximum grade, performance appraisal ratings, and awards earned. Personnel records have been used to categorize ATCSs as (1) successfully completing training at the first facility, (2) remaining in training status at the first facility, (3) being reassigned to another facility (at a lower grade), (4) being reassigned to another option (at a lower grade) before completing training, (5) separating from the series for reasons related to performance, and (6) separating from the series for other reasons. Those separating for reasons unrelated to performance are typically excluded from all analyses.

Job performance criteria

At the time this book was published, no measures of on-the-job performance were systematically obtained for use as selection test validation criteria by the FAA. Performance appraisal ratings are available from CPMIS, but have little variability and encompass areas besides technical job performance. Some believe that operational errors or deviations should be considered an "ultimate" criterion because of their role as the product of the air traffic control system and their potential consequences. However, a few problems are associated with the use of operational errors. First, commission of an operational error is such a rare event that there should be little variability in individual scores. Second, operational errors may occur for a variety of reasons, which may not be described fully for an observer. It is sometimes difficult to determine a cause for an operational error because the method for attributing causation is not very precise. The causal factor categorization used in the final operational error report (Form 7210-3) is also not very descriptive. For these reasons, operational errors have not been utilized to date as criteria, although if the methods for reporting and describing operational errors are improved, they might be considered more relevant.

Alternative approaches to the measurement of ATCS performance

Simulations

Because of the criticality of air traffic control operations, it is often not feasible to obtain performance measures while controllers are working. One way to avoid this problem is to measure the performance of controllers working in simulated environments, as suggested 40 years ago by Taylor (1952). One question relevant to the use of criteria obtained from simulation devices relates to their fidelity. The concept of "fidelity" encompasses both system fidelity, that is, the match between the system used in the test and the system used operationally, and environmental fidelity (Meister, 1986), that is, the match between the environmental context used during the test and the environmental context typically present in day-to-day operations. Simulation devices can be configured

to recreate system and environmental fidelity as closely as possible, given the resources available to produce the desired amount of fidelity.

If the fidelity of the simulation is not perceived as closely resembling actual operations, the results may be perceived as inaccurate. Scenarios incorporated into simulations used as criterion measures should be constructed to accurately represent the complexity and type of activities found in the operational environment. Regardless of the realism built into a scenario, the knowledge that it is not real could affect the behavior of participating SMEs, as could the knowledge that their performance is being observed. Those who know they are participating in simulations might be expected to provide measures of maximum, rather than typical performance (Sackett, Zedeck, & Fogli, 1988). Observer-provided subjective ratings of SME performance could be influenced by different types of rater bias.

One problem specific to air traffic control involves decisions that will maximize generalizability to other situations. On one hand, controller performance can be affected by different attributes present in air traffic sectors, such as altitude, number of intersections, amount of traffic, presence of airports, traffic flow patterns, etc. With our present knowledge, we cannot determine whether 2 sectors are equally difficult for a controller. This factor complicates the ability to generalize findings obtained from one sector to another sector. If a controller working a specific sector obtains a criterion score and a controller working another sector obtains a different criterion score, how can the scores of the controllers be compared?

One solution to this problem might be to design a generic sector, or set of sectors, that all controllers work in order to obtain comparable scores. While a single sector could probably not be developed to encompass all the important properties on which sectors differ, a set of sectors might be developed to describe most generic situations that controllers encounter. However, use of such generic simulations poses another problem. Operational controllers develop extensive expertise by working in their airspace for many years. To what extent can the familiarity, experience, and expertise of operational experience be duplicated by working on a generic sector for a few hours? We do not know enough about how controllers develop expertise in their own airspace to be able to determine whether criterion measures obtained from a generic sector would be comparable to criterion measures obtained on the airspace with which the controller is familiar.

Buckley, DeBaryshe, Hitchner, and Kohn (1983) conducted 2 studies in which individual controllers ran air traffic problems in a generic sector under simulated conditions. The purpose of the study was to derive measures that would describe system functioning during simulations. Relative positions of aircraft, numbers and types of control actions taken, numbers and types of communications, and timing of events were recorded. Buckley et al. found that 4 factors adequately summarized the experimental results: (a) conflict (number of conflicts), (b) occupancy (time under control), (c) communications (duration of air-to-ground communications), and (d) delay (total delay time, number of aircraft handled, and fuel consumption). They also found that sector geometry interacted with number of aircraft to influence the results, suggesting that the sector configuration has an impact on performance measurement. Buckley et al.'s (1983) study, while exploratory, took the initial step towards measuring controller performance more objectively. Although their measures had multiple correlations between .58 and .74 with subjective performance ratings provided by observers, more work must be done to investigate the relationship between the objective measures and other measures of controller performance. Meister (1987) cautioned that it is necessary to understand interrelationships between system variables to assure that specific measures are directly related to operator performance.

Operational data replay and analysis: SATORI

A recent advance may soon allow controller performance to be measured using operational data. Situation Assessment Through the Recreation of Incidents (SATORI; Rodgers & Duke, 1993) is a system that allows for the re-creation of recorded air traffic data sent to a controller's Plan View Display (PVD) and Continuous Readout Update Display (CRD) for any sector. SATORI allows the re-creation of all ATC data available to be displayed to a controller for a given

period of time, including maps, aircraft positions and movements, weather, system messages, and controller inputs. These data are time-synchronized with recorded voice communications between the controller and pilots under the control of that sector. Thus, the system allows re-creating ATC events. The SATORI system was originally designed to re-create operational errors for review by quality assurance teams and controllers, and may be a useful tool for investigating other aspects of the interaction between controller, airspace architecture and complexity, and traffic load and complexity. For example, efforts are currently underway to customize the system to replicate the findings of Buckley, DeBaryshe, Hitchner, and Kohn (1983) using operational data. The objective measures derived from SATORI will be factor analyzed, along with observers' subjective reports of controller performance and the field training performance measures described above.

FUTURE DIRECTIONS

Air traffic control is such a dynamic job with such profound implications for aviation safety that it is important to develop performance criteria that adequately reflect its complexity. In earlier years, supervisor or instructor ratings of performance in training were used as validation criteria. Later, available measures of performance in selection programs (the ATC Screen) or in field training were used. These measures had certain desirable distributional qualities, but their relationship with job performance was questionable. More recent efforts have targeted looking at the product produced by the controller, using computer-generated scoring methods. While these types of measures may appear relevant, care must be taken to determine their relative importance and to what extent they are influenced by factors not under the control of the ATCS.

We return to the question of what we are trying to measure. In previous years, it appears that our criteria have been chosen for reasons of convenience, rather than relevance. As new methods to obtain criterion measures are developed, both managers and researchers must be careful to let the construct to be measured drive the choice of criteria. An optimal approach would involve the use of multiple types of criterion measures (e.g., objective, personnel, and judgmental indices; Landy & Farr, 1980). Similarly, the choice of predictor domains to be included in a test battery should be linked to a clear understanding of what aspects of job performance are to be predicted and the worker characteristics required to achieve the behaviors valued by the organization. For example, personality and biodata may have more relevance to what Borman and Motowidlo (1993) term contextual performance, while cognitive abilities may be more predictive of core technical criteria. Recent research in the U. S. by Schroeder, Broach, and Young (1992) suggests that non-cognitive predictors may have both incremental validity and financial utility in controller selection. The validation process suggested by Binning and Barrett (1989) offers a framework for developing models of human work behavior, encompassing multiple predictor domains, such as biodata and cognitive ability in relation to clearly articulated core technical and contextual performance criteria.

Sharing ATCS selection research results across nations would do much to further research describing the ATC job, defining and measuring job performance to be predicted, and developing tests to represent the worker characteristics required to safely and efficiently control air traffic. Such a pooling of research results will require that researchers provide more detailed information in published reports, when possible. For example, names of constructs are reported, without operational definitions or reference to standard taxonomies of human abilities. Tests used to represent constructs should be fully described, and example items given. Operational definitions of criteria and example data collection instruments would be very useful to researchers around the world. Correlation matrices, with predictor and criterion means, standard deviations, ranges, and sample sizes, should be presented in a table whenever possible. Such full reporting would enable researchers to match test constructs and conduct meta-analyses in order to identify commonalties and differences in ATC requirements. Only then might it be possible to begin to develop an international controller selection research program in support of an increasingly interconnected, global air traffic control system.

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. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA238267.

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* (pp. 7 - 12). (DOT/FAA/AM-91/9). Washington, DC. Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA238267.

Binning, J.F., & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74,* 478 - 494.

Boone, J.O. (1984). Validation of new ATCS selection tests on trainee and controller populations Part III 1976 - 1978 selection study of new appointees to the ATC occupation. In S.B. Sells, J.T. Dailey, & E.W. Pickrel (Eds.), *Selection of air traffic controllers* (pp. 375 - 395). (DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA147765.

Borman, W.C., & Motowidlo, S.J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, W.C. Borman, & Associates (Eds.), *Personnel selection in organizations,* pp. 71 - 98. San Francisco: Jossey-Bass.

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

Broach, D., & Manning, C.A. (1994). *Validity of the air traffic control specialist nonradar screen as a predictor of performance in radar-based air traffic control training.* (DOT/FAA/AM-94/9). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA279745.

Brokaw, L.D. (1984). Early research on controller selection, 1941-1963. In S.B. Sells, J.T. Dailey, & E.W. Pickrel (Eds.), *Selection of air traffic controllers* (pp. 39 - 77). (DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA147765.

Browne, D.S. (1993, March). *Training of student air traffic controllers in the UK CAA.* United Kingdom: CAA House.

Buckley, E.P., & Beebe, T. (1972). *The development of a motion picture measurement for aptitude for air traffic control.* (FAA/RD-71/106). Atlantic City, NJ: Federal Aviation Administration Technical Center.

Buckley, E.P., DeBaryshe, B.D., Hitchner, N., & Kohn, P. (1983). *Methods and measurements in real-time air traffic control system simulation.* (DOT/FAA/CT-83/26). Atlantic City, NJ: Federal Aviation Administration Technical Center.

Campbell, C.H., Ford, P., Rumsey, M.G., Pulakos, E.D., Borman, W.C., Felker, D.B., De Vera, M.V., & Riegelhaupt, B.J. (1990). Development of multiple job performance measures in a representative sample of jobs. *Personnel Psychology, 43,* 277 - 300.

Campbell, J.P., McHenry, J.J., & Wise, L.L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43,* 313 - 333.

Cobb, B.B., Matthews, J.J., & Nelson, P.L. (1972). *Attrition-retention rates of air traffic controller trainees recruited during 1960-1963 and 1968-1970.* (DOT/FAA/AM-72/33). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA757933.

Collins, W.E., Boone, J.O., & VanDeventer, A.D. (1984). The selection of air traffic control specialists: Contributions by the Civil Aeromedical Institute. In S.B. Sells, J.T. Dailey, & E.W. Pickrel (Eds.), *Selection of air traffic controllers* (pp. 79 - 112). (DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA147765.

Dailey, J.T., & Pickrel, E.W. (1984). Development of the Multiplex Controller Aptitude Test. In S.B. Sells, J.T. Dailey, & E.W. Pickrel (Eds.), *Selection of air traffic controllers* (pp. 281 - 298). (DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA147765.

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. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA230058.

Eißfeldt, H. (1991). DLR selection of air traffic control applicants: Predictive validity. In E.Farmer (Ed.). *Human resource management in aviation: Proceedings of the XVIII WEAAP Conference, Volume I,* (pp. 37 - 44). Aldershot, UK: Avebury Technical.

Ekstrom, R.B., French, J.W., & Harman, H.H. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Federal Aviation Administration. (1985). *National air traffic training tracking system.* (FAA order 3120.22A). Washington, DC: Author.

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

Goeters, K.-M., Timmerman, B., & Maschke, P. (1993). The construction of personality questionnaires for selection of aviation personnel. *International Journal of Aviation Psychology, 3,* 123 - 141.

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.

Haglund, R. (1993). Swedish research program concerning selection and training of air traffic controller students. *Journal of ATC, 35*(3), 63 - 64.

Haglund, R., Andersson, B., Backman, B., Brehmer, B., & Sundin, O. (1993). *Recruitment, selection, and training of air traffic controller students: Summary Phase I.* (Delrapport Nr. 10). Norrkoping, Sweden: Swedish Civil Aviation Authority Air Navigation Service.

Hattig, H.-J. (1991). Selection of air traffic control cadets. In R.Gal & A.D. Mangelsdorff (Eds.), *Handbook of military psychology* (pp. 115 - 129). New York: Wiley.

Henry, J.H., Kamrass, M.E., Orlansky, J., Rowan, T.C., String, J., & Reichenbach, R.E. (1975). *Training of U.S. air traffic controllers.* (Institute for Defense Analysis report R-206). Washington, DC: Department of Transportation Federal Aviation Administration. [NTIS AD A006 603].

Hilton, T.F., & Dolgin, D.L. (1991). Pilot selection in the military of the free world. In R. Gal & A.D. Mangelsdorff (Eds.), *Handbook of military psychology* (pp.81 - 101). New York: Wiley.

Hilton, T.F., & Sells, S.B. (1984). Air traffic controller selection in the United States and other countries: An international overview. In S.B. Sells, J.T. Dailey, & E.W. Pickrel (Eds.), *Selection of air traffic controllers* (pp. 26 - 37). (DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA147765.

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..

Landy, F.J., & Farr, J.L. (1980). Performance rating. *Psychological Bulletin, 87,* 72 - 107.

Lewis, M.A. (1978). *Use of Occupational Knowledge Test to assign extra credit in selection of air traffic controllers.* (DOT/FAA/AM-78/7). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA05367/5GI.

Manning, C.A. (1991a). Procedures for selection of Air Traffic Control Specialists. In Wing, H. and Manning, C.A. (Eds.) *Selection of Air Traffic Controllers: Complexity, Requirements, and Public Interest* (pp. 13 - 22). (DOT/FAA/AM-91/9). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA238267.

Manning, C.A. (1991b). Individual differences in air traffic control specialist training performance. *Journal of the Washington Academy of Sciences, 81*, 101-109.

Manning, C.A. and Aul, J.C. (1992). *Evaluation of an alternative method for hiring air traffic control specialists with prior military experience.* (DOT/FAA/AM-92/5). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA246587.

Manning, C.A., & Broach, D. (1992). *Validity of the air traffic control specialist nonradar screen as a predictor of performance in radar-based air traffic control training.* Paper presented at the 4th Annual Convention of the American Psychological Society, San Diego, CA.

Manning, C., Della Rocco, P., and Bryant, K. (1989). *Prediction of success in FAA air traffic control field training as a function of selection and screening test performance.* (DOT/FAA/AM-89/6). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA209327.

Mecham, R.C., McCormick, E.J., & Jeanneret, P.R. (1977). *Position Analysis Questionnaire Users Manual (System II).* West Lafayette, IN: Purdue University.

Meister, D. (1986). Human factors testing and evaluation. In G. Salvendy (Ed.), *Advances in Human Factors/Ergonomics* (Vol 5.). Amsterdam: Elsevier Science Publishers.

Meister, D. (1987). System effectiveness testing. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 1271-1297). New York: Wiley.

Mies, J.M., Colmen, J.G., & Domenech, O. (1984). Validation of new ATCS selection tests on trainee and controller populations Part II 1977: Selection of air traffic control specialists. In S.B. Sells, J.T. Dailey, & E.W. Pickrel (Eds.), *Selection of air traffic controllers* (pp. 357 - 375). (DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA147765.

Milne, A. & Colmen, J.G. (1984). Validation of new ATCS selection tests on trainee and controller populations Part I 1972: Selection of air traffic control specialists. In S.B. Sells, J.T. Dailey, & E.W. Pickrel (Eds.), *Selection of air traffic controllers* (pp. 353 - 357). (DOT/FAA/AM-84/2). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA147765.

Mogford, R.H., & Tansley, B.W. (1992). *The importance of the air traffic controller's picture.* Ottawa, Canada: Hickling Corporation.

Nyfield, G.R., Kandola, R.S., & Saville, P.F. (1983, December). *The selection of air traffic controllers: Concurrent validity study.* Surrey, Kent, UK: Saville & Holdsworth, Ltd.

Robertson, I.T. & Kinder, A. (1993). Personality and job competencies: The criterion-related validity of some personality variables. *Journal of Occupational and Organizational Psychology, 66*, 225 - 244.

Rock, D.B., Dailey, J.T., Ozur, H., Boone, J.O., & Pickrel, E.W. (1981). *Selection of applicants for the Air Traffic Controller Occupation.* (DOT/FAA/AM-82/11). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA122795/8.

Rodgers, M. & Duke, D. (1993). *SATORI: Situation assessment through the re-creation of incidents.* (DOT/FAA/AM-93/12). Washington, DC: Federal Aviation Administration Office of Aviation Medicine. Available from: National Technical Information Service, Springfield, VA 22161; order # ADA268390.

Rodriquez, E., Narayan, K., & O'Donnell, R.D. (1986). *An approach to testing for neurobehavioral effects of organic solvents.* Los Angeles, CA: Atlantic Richfield Company.

Sackett, P.R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum performance. *Journal of Applied Psychology, 73,* 482 - 486.

Saville & Holdsworth, Ltd. (1990). *Occupational personality questionnaire manual.* Esher, Surrey, United Kingdom: Author.

Saville, P. & Wilson, E. (1991). The reliability and the validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology, 64,* 219 - 238.

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

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, L., & Lane, I.M. (1982). *Personnel and organizational psychology.* Homewood, IL: Richard D. Irwin, Inc.

SPSS, Inc. (1988). *SPSS-X user's guide.* (3rd Ed.). Chicago, IL: Author.

Svensson, B., & Trygg, L. (1991). *Personality characteristics of candidates for air traffic controller training.* University of Lund, Sweden: Department of Applied Psychology.

Swinburne, P.J. (1985). A comparison of the OPQ and 16PF in relation to their occupational application. *Personnel Review, 14,* 29 - 33.

Taylor, M.V. (1952). *The development and validation of a series of aptitude tests for the selection of personnel for positions in the field of air traffic control.* Pittsburgh, PA: American Institutes for Research.

VanDeventer, A.D. (1981). *Field training performance of FAA Academy air traffic control graduates.* Paper presented at the Annual Scientific Meeting of the Aerospace Medical Association.

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.

Weick, K.E. (1979). *The social psychology of organizing.* New York: Random House.

☆U.S. GOVERNMENT PRINTING OFFICE: 1997 - 559-079/60008