FLIGHT SAFETY FOUNDATION
Airport Operations

Vol. 24 No. 2 *For Everyone Concerned with the Safety of Flight* March–April 1998

International Sharing of ATC Research Urged

A report by the U.S. Federal Aviation Administration Civil Aeromedical Institute examined air traffic controller selection in the United States, Germany, the United Kingdom and Sweden. The report also reviewed issues involved in air traffic controller job-performance measurement and found that questions remain about what performance criteria are relevant and how they are to be measured.

FSF Editorial Staff

A report published in July 1997 by the U.S. Federal Aviation Administration (FAA) Civil Aeromedical Institute (CAMI) examined the means of air traffic controller selection in the United States, Germany, the United Kingdom and Sweden.

Air traffic control (ATC) has long been recognized as a profession whose practitioners must have an unusual combination of work skills and personality traits. Measuring those skills and traits is of critical importance to the national civil aviation authorities (CAAs) that recruit and employ controllers.

The report said, "Sharing ATCS [air traffic control specialist] 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 reports 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. …

"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."

Many are called to ATC work, but few are chosen. In the 1982–1988 period, for example, only 7.4 percent of the applicants in Germany were selected for ATCS positions.

"Similarly," said the report, "11,280 persons out of 238,946 applicants in the United States successfully completed both the first-stage written aptitude examination and second-stage work sample at the FAA Academy [at the Mike Monroney Aeronautical Center, Oklahoma City, Oklahoma, U.S.] for a selection rate of about 5 percent during the same period."

Measuring the attributes of controllers — and controller candidates — is an evolving discipline, still far from an exact science.

In the United States between 1976 and 1992, ATCS selection involved a four-step process:

- A written aptitude test battery;
- A personal interview;
- A medical examination; and,
- Performance-based screening at the FAA Academy.

During that period, the written tests included the Multiplex Controller Aptitude Test (MCAT) and the Abstract Reasoning Test (ABSR).

The report said, "The MCAT was designed to measure applicants' skills in applying a simplified set of ATC rules within a simulated [ATC] environment. … 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."

Controllers were hired provisionally, pending further testing. They advanced to a series of performance-based, or "second-level" measurements, known as the ATCS screen.

"A total of 13 performance assessments, including classroom tests, laboratory simulations of nonradar [ATC] and a final written examination, were made during the course of the ATCS screen," said the report.

That elaborate selection process was costly to the FAA and to the applicants who were selected for the ATCS screen. The FAA spent US$20 million to $25 million per year to choose about 1,400 trainee controllers. And the provisional hires had to leave their previous jobs and sometimes their families for nine weeks during the ATCS screen. Fewer than two-thirds of them remained as controllers at the end of that assessment.

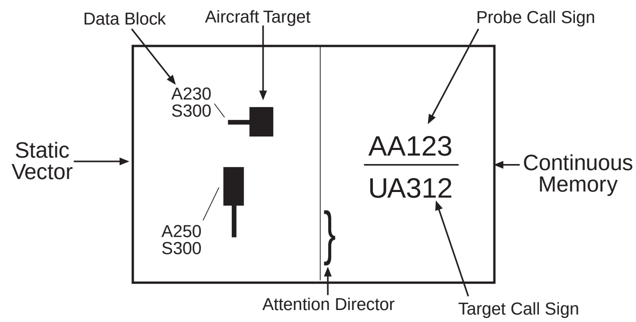
"That risk may have discouraged potentially qualified women and racial minorities from pursuing an air traffic career," said the report. "The FAA undertook a major review of its ATCS selection program in 1990 to address these costs and other concerns.

"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 nine-week ATCS screen … ." An interim computer-administered test battery was developed beginning in 1990 by studying what researchers had determined about the abilities needed for the ATCS job.

"U.S. 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," said the report. "Two computer-administered information-processing tests were designed to dynamically assess cognitive attributes, such as spatial reasoning, short-term memory, movement direction, pattern recognition and attention allocation."

Beginning in 1991, the FAA began experimenting with computerized information-processing tests. Those tests included the Static Vector/Continuous Memory (SV/CM) test (Figure 1) and the Time Wall/Pattern Recognition (TW/PR) test (Figure 2).

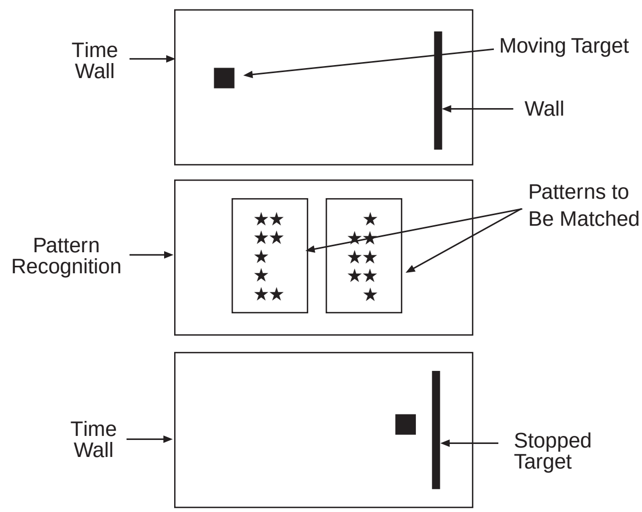
U.S. Federal Aviation Administration
Static Vector/Continuous Memory Screen



Source: U.S. Federal Aviation Administration Civil Aeromedical Institute

Figure 1

U.S. Federal Aviation Administration
Time Wall/Pattern Recognition Screens



Source: U.S. Federal Aviation Administration Civil Aeromedical Institute

Figure 2

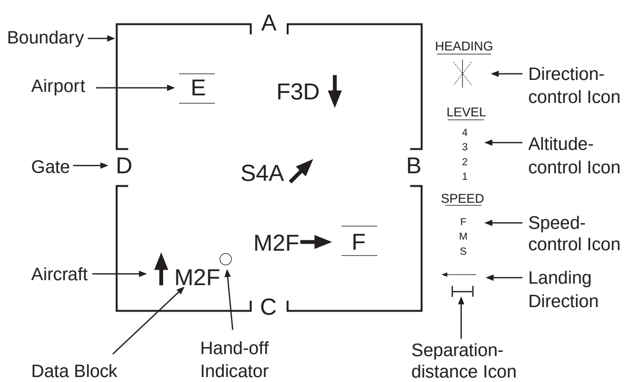
"The SV component requires the subject to make judgments about conflicts, while the CM component exercises working memory," said the report. "When the attention director [is] to the left, the subject's task [is] to decide if the aircraft targets [will] collide or not, based on the altitude ('A230') and speed ('S300') information in the data blocks and spatial relationships of the targets. When the attention director [is] to the right, the subject's task [is] to first, memorize the target call sign *below* the line, and second, indicate if the probe call sign *above* [is] the same [as], or different [from] the target call sign that had been presented *below* the line in the previous CM trial."

The TW/PR screen measures the subject's ability to estimate speed of movement and to determine whether patterns are the same or different.

"First," the report said, "the target appear[s], moving from left to right at a steady speed toward the 'wall' (top screen). After an initial time interval, the target and wall [are] masked by a pair of patterns (middle screen). The subject's task [is] to decide if the patterns [are] the same or different. A new pair of patterns appear[s] after each response [is] made. However, the subject ha[s] to keep in mind the continuing movement of the TW target toward the wall, as the TW task [is] to stop the target (bottom screen) as close as possible to, without actually hitting or passing through, the wall."

Besides the information-processing tests, applicants are assessed on a computerized, low-fidelity simulation of ATC vectoring and separation tasks, an exercise called the Air Traffic Scenario Test (ATST; Figure 3).

U.S. Federal Aviation Administration Air Traffic Scenario Test Screen



Source: U.S. Federal Aviation Administration Civil Aeromedical Institute

Figure 3

"The boundary encloses a simplified airspace, with four outbound gates, A, B, C and D, and two airports, E and F," said the report. "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."

The SV/CM, TW/PR and ATST together became known as the ATCS pretraining screen (ATCS/PTS), which was validated in two studies. One study was predictive, using newly hired ATC students as subjects and assessing how well the ATCS/PCS predicted performance in the ATCS screen (that is, the existing test methodology); the second was concurrent, correlating trainees' and full-performance-level controllers' scores on the ATCS screen, the ATCS/PCS and a composite score of various measures of the controllers' success in field training and on-the-job training (OJT).

Both validation studies indicated that the ATCS/PTS could replace the ATCS screen as the second phase in selecting ATCSs.

"The ATCS selection system," the report said, "now consists of the four-hour written ATCS aptitude test battery followed by, for those applicants earning a qualifying score, second-level screening on the ATCS/PTS."

The report discussed controller selection systems and supporting research in Germany, the United Kingdom and Sweden, based on documentation available to the authors.

The German Aerospace Research Establishment (DLR), which is in many activities analogous to the FAA Technical Center and CAMI, conducts controller-selection research.

"The DLR," said the report, "has developed and validated a four-step ATC selection procedure that requires about four days to administer:

"1. A 'preselection' phase consisting of a battery of eight 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; [and,]

"4. Part III of the 'main-selection' phase, consisting of an interview with a board comprised of a senior controller, two other experienced controllers and two DLR aviation psychologists as advisors."

The DLR selection procedure is based in part on measurement of candidates' cognitive and personality traits (Table 1, page 4).

The German ATCS selection tests have been validated against success in training, although not actual ATCS on-the-job performance. Validities were found to be significant between the selection tests and two written examinations, a radar simulation and final grades in training.

In the United Kingdom, the U.K. CAA, anticipating a greater need for controllers in the 1990s, undertook in 1982 a project that involved ATCS job analysis and a testing-validation study. The job analysis resulted in a model of ATCS capabilities and traits shown in Table 2, page 4.

Table 1
Traits Assessed in the German Air Traffic Control Specialist Selection System

Performance Domain	Personality Domain
Basic Knowledge	**Achievement-oriented Traits**
English	Motivation
Technical Comprehension	Rigidity
Mathematical-logical Thinking	Mobility
	Vitality
Operational Attitudes	**Interpersonal Behavior**
Memory	Extroversion
Perception and Attention	Dominance
Spatial Orientation (Auditory/Visual)	Aggressiveness
	Empathy
Multiple-task Capacity	
	Stress Resistance
	Emotional Stability

Source: U.S. Federal Aviation Administration Civil Aeromedical Institute

"The core of controller skill appears to involve rapid processing of information from multiple channels in order to develop and maintain a 'real-time' representation of events in the airspace," said the report. "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."

Six tests were developed to evaluate abilities associated with controller skills.

The report said:

- "In the 10-minute **basic checking test**, the examinee [is] required to find the number or letter [text] string from among five alternative strings on the right-hand page that exactly [match] the probe string on the left-hand page;

- "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 five lines, angles or figures [is] presented to the examinee in each item. The examinee's task is to identify the two lines, angles or figures that are identical.

- "The **spatial reasoning test** (20 minutes) presents a pattern, which, when folded, creates a cube. ... 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 three dimensions mentally).

- "The **diagramming test** (20 minutes) is described as measuring 'logical analysis through the ability to follow complex instructions.' The stimulus consists of one 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 five alternatives, the column of boxes resulting from carrying out the operations

Table 2
United Kingdom Model of the Air Traffic Control Specialist 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 nonroutine	Decisiveness and confidence
	Checking/updating information	Conscientiousness
	Short-cycle repetitive work	Structured thinking
		Self-control

Source: U.S. Federal Aviation Administration Civil Aeromedical Institute described by the stimulus. This test may represent a measure of nonverbal general reasoning ability ... .

- "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 nonverbal general reasoning ability."

The ATCS job analysis commissioned by the U.K. CAA also resulted in a personality test being included in the test battery. The Occupational Personality Questionnaire (OPQ) was designed to assess an examinee on 32 personality traits, each of which falls within the "relating domain," "thinking domain" or "feeling domain" (Table 3).

Table 3
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
Effusive	Innovative	Possessing self-esteem
Tolerant	Tolerant of ambiguity	Active
Gregarious	Artistic	
	Decisive	

Source: U.S. Federal Aviation Administration Civil Aeromedical Institute

"The test battery was implemented in the mid-1980s, and is the subject of continuing research and evaluation," said the report. "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 [between test results and success in training] reported by [researchers is] 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 ... ."

In Sweden, ATCS selection and training are administered by the Civil Aviation Administration (LFV). The selection involves a series of tests and interviews.

The report said, "The aptitude tests used by the LFV were grouped into four general factors [Table 4, page 6]: (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 LFV recently began to formulate the basis for a revised ATCS selection system. The validity of the current tests was examined and the Swedish ATC system was analyzed to determine what traits were most important in successfully performing the work.

The validities of the individual components of the tests were analyzed using training outcome as the criterion. The greatest correlations between passing the training course and test scores was found for the "ship's destination" and "proofreading" tests.

Interviews of 127 incumbent ATCSs provided the material for a questionnaire concerning the behaviors used by skilled controllers to cope with stressful situations or events. The questionnaires were tailored for each of three operational environments — tower, approach control and en route control. A representative sample of 158 ATCSs and instructors rated, on a seven-point scale, the relative importance of each coping behavior and the frequency of each stressful situation or event in the course of their daily work.

The report said, "The stressful situations or events were grouped into five 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 five categories: (1) decision making; (2) self-confidence; (3) information gathering and processing; (4) social relations; and (5) communications.

"'Information gathering and processing' behaviors were most effective in relation to coordination and traffic processing events in en route control towers. 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."

The report said, "A critical issue ... in the development and validation of a new generation of tests for all countries is measurement of controller job performance." The report considered how ATCS job performance is to be measured, and derived its discussion from the systems used, in the past or the present, in the United States.

"Measuring ATCS performance is critical [because] ... 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," said the report. "Controller selection research has relied upon training success as *the* criterion for validation of tests, rather than job performance."

The successful, or unsuccessful, completion of ATCS training lasting one year to three years has often been used as a validation criterion for the selection tests that determine who enters training. But success in training is not the ultimate goal of selection.

The report said, "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 [ATC]."

Another requirement for performance measurement is that it differentiate between typical and maximum performance.

"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?" said the report.

Predicting the success of ATCS trainees was historically based largely on supervisors' assessments. In the 1970s through the early 1980s, measures based on the ATCS screen supplemented supervisors' ratings.

The report said, "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.

"[Despite] their apparent advantages, several problems were associated with the use of these scores as test-validation criteria. First, because the [FAA] 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 [laboratory] problems at the time of testing; thus, performance was measured somewhere along the learning curve rather than at asymptote [the point at which the curve flattens out — that is, when learning is essentially complete].

"Second, the scores obtained for the laboratory problems were based on two 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."

During the 1980s and 1990s, performance measures based on radar training and OJT were used as predictive criteria. Information collected about trainees' OJT performance, including start and completion dates of training, number of hours taken to complete OJT, the trainees' grades and instructors' or supervisors' ratings of trainees' potential, was also used as a predictive criterion.

"As with the screen and radar-training measures, these field-training measures do not represent FPL [full-performance level] job performance," said the report. "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 the need to use the controller in an operational position, the number of other

Table 4
Swedish Civil Aviation Administration Selection Test Battery

General Factor	Description	Test
Flexibility and ability to find new solutions	Ability to improve and make decisions in unexpected situations	Ship's Destination Instruction Test II Necktie
Logical ability	Logical ability	Raven's Progressive Matrices Raven's Number Series
Spatial ability	Ability to construct a three-dimensional picture of the airspace from two-dimensional information	Blocks Metal Sheet Models Puzzles
Attention to detail, carefulness, short-term memory	Attention to detail, carefulness, short-term memory	Proofreading Number proofreading Name memory Number memory Figure identification

Source: U.S. Federal Aviation Administration Civil Aeromedical Institute 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."

Personnel records showing "career progression" have been used by the FAA to categorize ATCSs. Such records reflect an ATCS's status 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 (at a lower grade) before completing training; (5) separating from training for reasons related to performance; or (6) separating for other reasons.

No measures of on-the-job performance have been systematically obtained by the FAA to validate test criteria. Operational errors are believed by some to be the ultimate criterion for job performance because of the severity of their potential consequences. But the report noted that there are problems with using this measure as well.

"First, commission of an operational error is such a rare event that there should be little variability in individual scores," said the report. "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."

Because measuring ATCSs' performance while they are actually working might be a safety hazard, simulation of ATC situations offers an attractive alternative.

"One question relevant to the use of criteria obtained from simulation devices relates to their fidelity," said the report. "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 — that is, the match between the environmental context used during the test and the environmental context typically present in day-to-day operations."

A simulation environment must closely resemble the complexity and activity found in an actual ATC environment, but even if it does, an ATCS cannot fail to know that it is a simulation and that he or she is being observed.

The report said, "Those who know they are participating in simulations might be expected to provide measures of maximum, rather than typical, performance."

To what extent controller criterion scores in a specific simulated ATC sector can be generalized to represent performance in any sector is an unresolved question. Sectors can significantly vary in volume of traffic, altitudes of traffic, number of intersections, presence of airports and traffic-flow patterns. If two controllers receive similar criterion scores, but in different sectors, it is not clear whether the similarity of scores indicates similar abilities in both controllers.

"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," said the report. "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."

One recent development in measurement of ATC operations is situation assessment through the re-creation of incidents (SATORI). Using SATORI software, recorded air traffic data from a controller's plan view display (PVD) and continuous readout update display (CRD) can be replayed and synchronized with recorded voice communications between the controller and pilots in the controller's sector.

The report said, "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."

In a glance at future directions in controller selection and testing, the report said, "We return to the question of what we are trying to measure. 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.

"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."♦

Editorial note: This report was based on *Review of Air Traffic Controller Selection: An International Perspective*, U.S. Federal Aviation Administration Civil Aeromedical Institute, Report no. DOT/FAA/AM-97-15. July 1997. The 24-page report, written by Dana Broach and Carol A. Manning, contains figures and tables.

Disaster Response Planning Workshop for Business Aviation

June 18–19, 1998
Atlanta Airport Hilton and Towers
Atlanta, Georgia, U.S.

Who Should Attend?

- Department managers (flight, maintenance, scheduling and administration);
- Flight safety managers;
- Corporate safety/disaster response managers;
- Corporate security managers;
- Human resource/personnel managers;
- Public relations/communications managers;
- Risk/insurance and financial managers; and,
- Administrative managers.

Why Should You Attend?

- Develop your own disaster response plan—now!;
- Update your current disaster response plan (at least every other year);
- Increase the number of people in your department with skills and expertise in disaster response (one or two aren't enough);
- Improve corporate managers' understanding of the unique issues involved in an aviation-related disaster (you'll want all the help you can get); and,
- Help your department's staff after a nonaviation disaster (automobile accident, fire or act of violence).

Presented by

Flight Safety Foundation

The VanAllen Group, Inc.

For more information, contact: Joan Perrin, Flight Safety Foundation
Telephone: (703) 739-6700 • Fax: (703) 739-6708

Visit our World Wide Web site at http://www.flightsafety.org

AIRPORT OPERATIONS

ISSN 1057-5537

Subscriptions: US$60 (U.S.-Canada-Mexico), US$65 Air Mail (all other countries), six issues yearly. • Include old and new addresses when requesting address change. • Flight Safety Foundation, 601 Madison Street, Suite 300, Alexandria, VA 22314 U.S. • Telephone: (703) 739-6700 • Fax: (703) 739-6708

What's Your Input?

In keeping with FSF's independent and nonpartisan mission to disseminate objective safety information, Foundation publications solicit credible contributions that foster thought-provoking discussion of aviation safety issues. If you have an article proposal, a completed manuscript or a technical paper that may be appropriate for *Airport Operations,* please contact the director of publications. Reasonable care will be taken in handling a manuscript, but Flight Safety Foundation assumes no responsibility for submitted material. The publications staff reserves the right to edit all published submissions. The Foundation buys all rights to manuscripts and payment is made to authors upon publication. Contact the Publications Department for more information.