DOT/FAA/AM-03/20

Office of Aerospace Medicine
Washington, DC 20591

Use of Personality Assessment Measures in the Selection of Air Traffic Control Specialists

Raymond E. King[1]
Paul D. Retzlaff[2]
Cristy A. Detwiler[1]
David J. Schroeder[1]
Dana Broach[1]

[1]Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

[2]University of Northern Colorado
Greeley, CO 80639

December 2003

Final Report

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

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-03/20		

4. Title and Subtitle	5. Report Date
Use of Personality Assessment Measures in the Selection of Air Traffic Control Specialists	December 2003
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
King RE[1], Retzlaff PD[2], Detwiler CA[1], Schreoder DJ[1], Broach D[1]	

9. Performing Organization Name and Address		10. Work Unit No. (TRAIS)
[1]FAA Civil Aerospace Medical Institute P. O. Box 25082 Oklahoma City, OK 73125	[2]University of Northern Colorado Psychology Department Greeley, CO 80639	11. Contract or Grant No.

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

15. Supplemental Notes

Work was accomplished under approved subtask AM-B-03-HRR-523.

16. Abstract

Three studies illustrate the uses of personality assessment methods for selection of Federal Aviation Administration (FAA) air traffic control specialists (ATCSs). Study 1, using a select-out approach to the problems of screening, compared the Sixteen Personality Factor Questionnaire (16 PF) results with NEO Personality Inventory-Revised (NEO PI-R) results from 122 student ATCS participants. Results suggest that the current approach to personality assessment during the initial medical examination focuses primarily on the extent to which the applicant reports symptoms consistent with neurotic, inefficient, and perhaps argumentative characteristics. The remaining two studies consider select-in strategies looking at the psychometric issues of reliability, specificity, and validity. Study 2 examined the NEO PI-R along with the Experiences Questionnaire (EQ) subtests of the Air Traffic-Selection and Training (AT-SAT) battery scores. Participants included 142 students in the ATCS training program at the Mike Monroney Aeronautical Center, Oklahoma City, OK. That sample, in combination with an additional 99 participants who took only AT-SAT, allowed for the examination of the characteristics of the EQ itself. Scores on the eight subtests of the EQ were found to be moderately to highly intercorrelated, ranging from .53 to .81. A principal component analysis revealed only one major underlying factor for the EQ, accounting for 69.22 % of the variance. The overall EQ correlated most notably with the NEO PI-R domains of *Neuroticism* and *Conscientiousness*. Finally, Study 3 examined one of the more recent personality measures to be used for personnel research, the Armstrong Laboratory Aviation Personality Survey (ALAPS). A total of 121 students at the FAA Academy completed both the ALAPS and AT-SAT. Scores on the ALAPS *Depression* scale were negatively correlated with scores on the composite AT-SAT, and scores on the *Organization* scale were positively correlated with scores on AT-SAT, suggesting that there may be dimensions of the ALAPS that are related to the overall skills and abilities required for individuals to achieve success as ATCSs. This review affords the reader an appreciation of the uses and history of personality assessment, both select-out and select-in, in FAA ATCS selection. The FAA is continually evaluating cognitive, as well as personality, measures as part of its selection procedures.

17. Key Words	18. Distribution Statement
Personnel Selection, Personality, Personality Tests, Psychological Tests, Psychometrics, Screening, Selection	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	15	

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

ACKNOWLEDGMENTS

This technical report could not have been completed without the generous assistance of Roni Anderson, Edna R. Fiedler, David Hunter, Kai-Peter Koenig, Gene K. Nebel, Steven Schwendeman, Katherine Wade, and Mike Wayda.

Use of Personality Assessment Measures in the Selection of Air Traffic Control Specialists

The conduct of psychological testing programs to identify the most promising candidates for Federal Aviation Administration (FAA) air traffic control training is over half a century old. The year 1952 saw publication of a report by the American Institute for Research entitled *The development and validation of aptitude tests for the selection of personnel for positions in the field of air traffic control* (Taylor, 1952). Air traffic control personnel desired a screening battery of psychological tests for replacing the work force in the event of a sudden loss of personnel due to the exigencies of a major war. This project commenced with a job analysis. With the possible exception of the traits "carefulness" and "judgment," personality characteristics were not identified as being of primary interest. In February 1956, personnel from the Air Traffic Control Branch, Federal Airways Standardization Division, Civil Aeronautical Center, Oklahoma City, OK, visited the USAF Personnel Laboratory at Lackland AFB (San Antonio), TX, to learn about the Air Force's methods of selecting control tower and air traffic control personnel (Selection and Classification Branch, 1958). Brokaw (1959) described the various tests studied and reported that "temperament measures fail to demonstrate useful validity for either academic or job criteria" (p. 6). However, he further reported that of the 12 California Test of Personality scales, only the *Family Relations* scale was predictive of instructor rating. Moreover, scores on the *Nervous Manifestations* scale from the California Test Bureau Mental Health Analysis predicted academic grades, while these scores were unrelated to instructor or supervisory ratings, or promotion criteria. Simultaneously, the Flight Safety Foundation published a report (Kraft, 1958) under contract with the Civil Aeronautics Administration, forerunner of the FAA, forming the basis for what would become the Health Program for Agency Air Traffic Control Specialist Employees in 1965 (FAA Order 9430.2). This order set the medical and psychological standards for entry into, and continuation in, the air traffic control occupation, as will be described shortly.

The first technical report published by the Civil Aeromedical Research Institute (CARI, forerunner of today's Civil Aerospace Medical Institute, CAMI) was entitled *Problems in air traffic management: I. Longitudinal prediction of effectiveness of air traffic controllers* (Trites, 1961). Trites found that psychological testing, including scales from a personality measure (California Test of Personality), was a potentially useful screening tool. Trites also found that psychological testing predicted later supervisor ratings. A subsequent study, published only months later, entitled *Problems in Air Traffic Management: II Prediction of success in Air Traffic Control School* (Cobb, 1962), addressed similar issues. Cobb found psychological testing to be a useful tool; however, measures derived from the California Test of Personality were less predictive of success in air traffic control training than were the other (more cognitive) tests examined.

Decades later, Collins, Schroeder, and Nye (1989), using the State-Trait Personality Inventory (STPI), found that scores on the STPI scales measuring anxiety were inversely related to successful training and good on-the-job performance. Another study (Schroeder, Broach, & Young, 1993), using the NEO Personality Inventory (NEO PI), found that successful air traffic control students exhibited lower levels of *Neuroticism*, higher average scores of *Extraversion*, *Openness to Experience*, and *Conscientiousness,* and no difference on *Agreeableness* when compared with a normative sample on the NEO PI. Schroeder et al. (1993) also found that the NEO PI (specifically the facets *Fantasy*, *Activity*, and *Ideas*) explained additional variance in predicting who would successfully complete the FAA Academy Screen program (a selection and training program). However, the overall contribution of these personality variables was very small.

In the mid-1960s, in an effort to monitor the health of air traffic control personnel, the FAA commenced initial and annual examination of all air traffic control specialists (ATCSs) under the Air Traffic Controller Health Program (ATCHP). "The purpose of the FAA's Air Traffic Controller Specialist (ATCS) Health Program is to help every controller stay in good health, to maximize the productive working life of ATCSs, and to maintain a safe and efficient air traffic system" (as outlined in FAA Order 9430.2). The ATCHP was intended to focus on vision, hearing, and the cardiovascular and nervous systems and called for use of medical and psychological screening examinations. The rationale was that ATCSs face "stressful working conditions and need decision-making skills" ("FAA Will Require Exams," 1966, p. 17). There was interest in testing the "*generally held but scientifically unsubstantiated belief that the stresses encountered in air traffic control work may, over a period of time, impair the health and personalities of air traffic controllers*." At stake were "*Early retirement and other career benefits*" (DOT, 1969, p. 95). The goals, as outlined in FAA Order 9430.2, *Establishment of Health*

Program for Agency Air Traffic Control Specialist Employees, 15 October 1965, included detection of early indications of health problems in controllers to:

- protect the health of the controller,
- determine their effectiveness in the job, and
- ensure the safety of the flying public.

During the time this FAA order was initially implemented, critics of personality testing for employment screening, including members of the Senate and the House of Representatives[1], the ACLU[2], and some governmental union officials[3], claimed invasion of privacy. Focusing on the individual items of the tests used at that time rather than on scales, they cited objectionable items regarding:

- religion,
- birth control,
- preference for a "well-made gun versus a beautiful poem," and
- dating habits at the age of 15 years.

On June 9, 1965, R. Sargent Shriver, Director of the Peace Corps, provided a notable contrast during a hearing before the Subcommittee on Constitutional Rights entitled *Psychological Testing Procedures and the Rights of Federal Employees* (June 7 - 10, 1965; Senate Report No. 56-310). Shriver testified that the program of psychological testing in the Peace Corps enjoyed apparent good results. Only 8% of those selected failed to complete their service for "personal adjustment" issues and less than .7% of those selected were returned for psychiatric difficulties.

Historically, Guion and Gottier (1965) discouraged the use of personality assessment in employee selection. With all the controversy about personality testing and about whether general intelligence is the best predictor of success in aviation careers (Carretta & Ree, 2003), why bother with personality testing? The emergence of the "five-factor" model (or "big-five" model, consisting of neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness) of personality, however, has resulted in a renewed interest in personality assessment in selection. Also, "Incremental validity over general mental ability appears to require measurement outside the strictly cognitive domain – e.g., psychomotor ability, social skills, perceptual speed, or personality" (Schmidt, Ones, & Hunter, 1992, p. 634). Finally, while cognitive tests yield high criterion-related validities, their predictive power may come with the price of adverse impact. Personality tests, however, can be used in combination with cognitive tests to provide measures that remain valid and reduce adverse impact (Dean, Russell, & Farmer, 2002). The matter is not straightforward; using personality data to reduce adverse impact on one minority group may increase adverse impact on another minority group (Ryan, Ployhart, & Friedel, 1998).

Employers value certain behaviors in employees. Individuals are expected to behave within certain parameters within the context of an organization. These parameters may be of varying widths; for example, compare the stereotypical view of a fighter pilot with that of an elementary school teacher. Some contend that the work environment itself provides the cues that drives an employee's behavior (Mischel, 1969), while others counter that personality is, by definition, the characteristic way of behaving and remains relatively constant throughout a person's life (Costa & McCrae, 1992). In other words, employees bring their personalities with them to work. Schneider, Smith, Taylor, and Fleenor (1998) proposed and empirically tested an attraction-selection-attrition model, which suggested that organizations are relatively homogenous in terms of the personality characteristics of their employees.

Barrick and Mount (1991), using the five-factor model, pointed to the central role of conscientiousness in predicting job performance. The five-factor model, however, is not without its critics, and even proponents may have differing definitions. Hough (1992) proposed a nine-factor model and identified "achievement" (which she characterized as the "tendency to strive for competence in one's work," "works hard," and "concentrates on, and persists in, completion of the task at hand") as playing a central role in job performance. The person with high achievement "is also confident, feels success from past undertakings, and expects to succeed in the future" (p. 144). These traits may serve as the best predictor of workplace performance. It is reasaonable to assume that achievement, however, is at least a component characteristic of conscientiousness. In any case, Hough also points out the necessity of identifying the demands of the job as they relate to the characteristics of the potential employee.

While employers may be somewhat unsure of what type of employee would best fit into their organizations, they are usually more certain of the types of employees they do not desire. Select-out criteria, or guidelines for eliminating applicants with a disqualifying psychiatric diagnosis (lack of fitness), results in the identification

[1] Representative Cornelius E. Gallagher, 28 April 65, on House floor and Senator Sam J. Ervin, Jr. to Najeeb E. Halaby, FAA Administrator, dated 28 May 65.

[2] In a letter to John W. Macy, Chairman of the U.S. Civil Service Commission, dated 4 May 66.

[3] *A Million $ Mistake — Punchcard Psychometry,* National Association of Government Employees, Local R12-5, Los Angeles ARTC Center, Palmdale, CA, 22 January 68.

of a very small subset of the candidate pool but does not identify the most qualified or adaptable applicant. Select-in methods determine who is best suited for challenging tasks but are relatively ineffective at screening for psychopathology.

No selection system is perfectly accurate; all will involve a degree of error. Errors in prediction can be of two types. A Type 1 error results when an applicant who *could* have been successful is rejected, and a Type 2 error results when an applicant is accepted and is ultimately unsuccessful. We rarely, if ever, have the opportunity to gauge the extent of our Type 1 errors, as the individual is rejected and therefore we never know how they *would* have performed. Pre-employment screening, or post-employment screening after a conditional offer of employment is tendered, is frequently used as a method to identify candidates for additional assessment, either from a select-out or select-in perspective.

Several factors must be considered before a test can be used with a population. First, is it reliable, or in other words, is the test score itself relatively free from error, internally consistent, and stable over time? Will an individual achieve a similar score on two separate testing occasions (assuming no learning or other change took place in the interim period between testings, including that due to the experience of the testing itself)?

Whether or not the test is valid for the population and question at hand is the next matter to consider. A test may be valid for one use and invalid for another. Validity is not an all-or-none proposition and involves inferences drawn from the test scores. A test may be valid to an extent in one situation and less valid in another situation. For example, an intelligence test would not be used to gauge citizenship. For a test to be considered valid for selection for a specific job, not only must its scales behave well internally, but the test must also actually differentiate on relevant criteria among candidates of known status. These issues are reviewed as a basis for consideration of the three empirical studies that follow. Because outcome data are lacking at the present time, our primary focus is on psychometric issues.

Study 1: Screening, Reliability, and Validity in Select-Out Testing

As stated in the introduction, ATCHP provides guidance on physical/medical certification requirements for ATCSs. The FAA uses the Sixteen Personality Factor Questionnaire (16PF; Cattell, Eber, & Tatsuoka, 1970) to assess the psychological fitness of individuals seeking entry into the air traffic profession. Since 1965, applicants have completed Forms A and B of the 16 PF as part of their medical examination. Commencing in 1978, the FAA has used a subset of 38 items (18 from Form A and 20 from Form B) as a "case finder" to identify individuals who must undergo additional assessment. Convey (1984) reported these items to be highly correlated with the second-order anxiety factor. In an unpublished report written in July 1996, Schwarzkopf, Buckley, and Pace from the University of Oklahoma noted the continually declining scientific interest in the 16 PF and the surging interest in the NEO-PI. Nevertheless, the authors commented favorably on the overall psychometric performance of the 16 PF 38-item scale described above. In the appendix, however, they noted some reservations about its use in high-risk occupations, its "fakeability," its sole focus on anxiety, and its low correlation with job performance measures.

The focus of Study 1 was to better understand the role of the 16 PF as a "case finder." In Study 1, we compared scores achieved using the existing 16 PF process with domain scores on the NEO-Personality Inventory-Revised (NEO PI-R; Costa & McCrae, 1992).

METHOD

Participants. One hundred twenty-two students in either terminal or en route training at the FAA Academy at the Mike Monroney Aeronautical Center in Oklahoma City, OK, served as voluntary study participants. Ninety men (74%) and 32 women (26%) participated. Their mean age was 26.5 years, with a range of 20 years to 54 years. The inclusion criterion was availability of both 16 PF results and NEO PI-R results.

Materials. The 16 PF, Form A and Form B (187 items each), and the NEO PI-R (240 items) were administered to each voluntary participant. The administration of these tests was not simultaneous, as applicants took the 16 PF as part of their medical examination and then were offered the opportunity to take the NEO PI-R when they came to the FAA Academy for ATCS training. The 16 PF was administered at the respective Regional Flight Surgeons' offices between 1999 and 2000. Sixteen PF items were presented with three response alternatives, while NEO-PI-R items were presented in a five-point Likert format from strongly disagree to strongly agree. For each of the keyed responses on the 18 items on Form A, the individual received one point. A total of at least 10 points is needed to clear the hurdle at this point. If this cut score was not achieved, then the 20 items from Form B were similarly scored, in a cumulative attempt to reach the cut score of 10. Interested readers are referred to Convey (1984) for more complete details on the FAA 16 PF scoring process. Of interest, three of the 18 items from Form A are scored counterintuitively; individuals receive a point for seemingly maladaptive responses.

RESULTS

On the basis of this two-tier scoring system, 26 (21%) of the 122 participants scored below 10 on the Form A items, necessitating scoring of Form B items. All 26 individuals passed the hurdle with the addition of the Form B items, thereby eliminating the need for further assessment for these individuals. The 18-item scale from Form A had a Cronbach alpha of .71, despite the counterintuitive scoring of three items. The 38-item scale (combining the 18 items from Form A with the 20 from Form B) had a Cronbach alpha of .85. The 18-item scale (on which a higher score equals better psychological health) was significantly correlated with several of the NEO PI-R domains (see Table 1).

Table 1. Correlations between the 18-Item Scale from the 16 PF (Form A) and NEO PI-R Domains

NEO PI-R Domain	Correlation
Neuroticism	-.44**
Extraversion	.11
Openness to Experience	-.02
Conscientiousness	.28**
Agreeableness	.23 *

* $p < .05$ (2-tailed)
** $p < .01$ (2-tailed)

DISCUSSION

In agreement with Schwarzkopf et al. (1996), the Cronbach alphas of the FAA 16 PF scales suggest relatively high internal consistency and suggest that the scales are reasonably reliable. Also in agreement with Schwarzkopf et al., the current screen-out approach appears to focus primarily on the extent to which the applicant reports symptoms consistent with neurotic, inefficient, and perhaps, argumentative characteristics. These traits, in the extreme, certainly seem consistent with mental health contraindication to the efficient performance of the duties of an ATCS. The current sample did not include any individuals who had been screened out as a result of their 16 PF results; those individuals would not have been given the opportunity to take the NEO PI-R. In the next study (Study 2), we examined the behavior of the NEO PI-R in a select-in context.

Study 2: Reliability, Specificity, and Validity in Select-In Testing With Current Tests

Traditionally, in an effort to better understand the dimensions that predict success in air traffic control training, CAMI scientists collect personality data, along with biographical ("biodata") questionnaires and cognitive abilities tests. These instruments are administered to students upon their entry into the ATCS training program at the FAA Academy.

Different jobs require differential attributes for success, especially jobs as highly specialized as that of an ATCS. Even Carretta and Ree (2003), while advocating the primacy of the *g* factor (general intelligence), allowed that personality and other qualities play important roles in successful training. In any case, to select those individuals who possess the abilities to perform well and succeed on a specific job, one must use a job task analysis to identify the skills and abilities needed for that specific job. The Separation and Control Hiring Assessment (SACHA) project defined the job tasks and identified the worker requirements for the ATCS occupation (Nickels, Bobko, Blair, Sands, & Tartak, 1995). In the non-cognitive realm, subject matter experts (SMEs) identified the worker requirements as outlined in Table 2.

The task, then, was to devise a reliable and valid method to measure those aspects of an applicant's personality relevant to the job at hand. As will be described, a non-cognitive test was created with a behavioral, as opposed to an opinion or internal-states, focus (Paullin, Houston, Bruskiewicz, & McKee, 1992). The test creators believed that it was important to measure traits, as they would be exhibited by persons who are not yet ATCSs.

The trait of conscientiousness has proved to be one of the more consistent predictors of subsequent success in a variety of occupations (Barrick & Mount, 1991). The SMEs independently recognized this situation by identifying "working cooperatively," "task closure/thoroughness," and "commitment to the job," all components of conscientiousness. In addition to their potential role in the initial selection process for certain safety-critical occupations, personality assessments are used as part of the medical/psychological clearance process to screen out individuals who appear to be unfit for the occupation. These assessments are conducted after a conditional offer of employment, in accordance with the Americans With Disabilities Act of 1990 (1991).

Current assessment of individual characteristics of personnel entering the ATCS profession is focused in two main areas. The first involves a new selection test battery, Air Traffic-Selection and Training (AT-SAT), and

Table 2. Noncognitive worker requirements as identified by SMEs (Nickels et al., 1995, pp. 70-71)

Interpersonal	
Professionalism	The ability to establish respect and confidence in your abilities among other controllers.
Working Cooperatively	The willingness to work with others to achieve a common goal. This includes a willingness to voluntarily assist another controller if the situation warrants.
Personal Tolerance	The ability to accommodate or deal with differences in personalities, criticisms, and interpersonal conflicts in the work environment.
Work/Effort	
Self-Esteem	Having a positive opinion/image of oneself.
Self-Confidence	A belief that you are the person for the job and knowing that your processes and decisions are correct.
Aggressiveness	The ability to take control of a situation-to reach out and take correct action.
Self-Awareness	An internal awareness of your actions and attitutdes. This includes knowing your limitations.
Attention to Detail	The ability to recognize and attend to the details of the job that others might overlook.
Task Closure/Thoroughness	The ability to continue an activity to completion through the coordination and inspection of work.
Decisiveness	The ability to make decisions in a timely manner.
Consistency	The ability to behave consistently at work (e.g., dealing with coworkers in a consistent manner; consistently using the correct phraseology).
Flexibility	The ability to adjust or adapt to changing situations or conditions.
Concentration	The ability to focus on job activites amid distractions for short periods of time.
Composure	Thinks clearly in stressful situations.
Tolerance for High Intensity Work Situations	The ability to perform effectively and think clearly during heavy work flow.
Motivation	The desire to motivate oneself through challenges on the job and to progress to a higher level of skill.
Commitment to the Job	The desire to be an ATCS and work hard to be successful.

the second is a battery of personality tests, to include the NEO PI-R, described in Study 1. AT-SAT was operationally implemented in June 2002. The battery is composed of eight subtests: *Dials* (*DI*), *Applied Math* (*AM*), *Scan* (*SC*), *Angles* (*AN*), *Letter Factory* (*LF*), *Air Traffic Scenarios* (*ATST*), *Analogies* (*AY*), and the *Experiences Questionnaire* (*EQ*). The battery requires an average of seven hours to complete, including breaks, and participants are allowed a maximum of eight hours to complete the entire battery. While seven of the eight subtests are cognitive, the EQ is a subtest comprised of 135 items that assess work-related attributes based on self-reported past experiences. This inventory is currently comprised of nine scales: *Composure*, *Consistency of Work Behavior, Working Cooperatively*, *Decisiveness, Self-Confidence*, *Interpersonal Tolerance*, *Execution*, *Task Closure*, and *Unlikely Virtues* (not currently included in the total EQ score). Applicants are asked to indicate their level of agreement with statements of past, mostly work, experiences on a five-point scale, from 1 (definitely true) to 5 (definitely false). The content domain of the

scales was driven directly by the constructs identified during the SACHA project (compare scale names with the job requirements delineated in Table 2).

A color brochure sent to prospective applicants offers, in addition to the above sample item, the following information:

Air traffic controllers must possess certain work-related attributes to perform their job well. The Experience Questionnaire determines whether candidates have these attributes by asking about past experience. There are no correct or incorrect answers. People will respond differently based on what is true for themselves.

The objective of Study 2 was two-fold. First, we were interested in the specificity of the EQ subtests to determine the degree to which they capture separate and specific dimensions. Second, the study was intended as a validity analysis to better understand the relationship of the EQ to the big five as measured by the NEO PI-R (Costa & McCrae, 1992).

METHOD

Participants. Two hundred forty-one (165 male, 76 female) voluntary participants took AT-SAT at the time of their entry into the FAA Academy ATCS training program at the Mike Monroney Aeronautical Center, Oklahoma City, OK. This sample was used to examine the characteristics of the EQ itself and had an age range of 18 to 40 years of age, with a mean of 26.6 years. A principal component analysis was conducted to understand the factor structure of the EQ. Of the 241 participants, 142 (94 male, 48 female) also voluntarily took the NEO-PI-R. Their ages ranged from 18 to 30, with a mean age of 24.

RESULTS

The eight subtest scores of the EQ had moderate to high intercorrelations, ranging from .53 to .81 (see Table 3). All of these correlations were significant at $p < .01$ (2-tailed).

The *Decisiveness* scale had the highest correlations with the other scales (.64 to .82). Principal component analysis revealed only one major underlying factor for the EQ, accounting for 69.22 % of the variance, despite the presence of eight scorable scales. This factor had an eigenvalue of 7.61, while the next component had an eigenvalue of only .65.

The overall EQ correlated most notably with *Neuroticism* —.34 ($p < .01$) and *Conscientiousness* .35 ($p < .01$). The only other significant correlation was with *Agreeableness* .19 ($p < .05$) (see Table 4). We did not attempt to correlate individual EQ subscores with the domain scores of the NEO-PI-R on account of the intercorrelations among EQ subscores.

DISCUSSION

The EQ apparently measures aspects of how individuals present themselves in terms of neuroticism and conscientiousness or, in other words, emotional stability and ability/willingness to follow rules and to be orderly.

Table 3. Number of items, weighted means and standard deviataions (SD), and intercorrelation of EQ Scales

Scale	No of items	Mean Score[4]	SD	1	2	3	4	5	6	7
1. Composure	12	1.17	.26							
2. Consistency of Work Behavior	9	4.41	.99	.58						
3. Working Cooperatively	9	2.35	.53	.72	.64					
4. Decisiveness	13	.40	.09	.75	.70	.82				
5. Self-Confidence	9	.71	.14	.67	.62	.74	.78			
6. Interpersonal Tolerance	10	.34	.07	.59	.65	.56	.64	.54		
7. Execution	10	.37	.07	.61	.59	.71	.74	.71	.53	
8. Task Closure	12	.22	.05	.66	.77	.77	.81	.71	.64	.74

[4] These weighted scores are included here for future meta-analysis purposes. Future researchers should bear in mind that these scores represent the original weighing scheme.

While these traits may be useful indicators of future on-the-job success, it is possible that positive impression management, or an effort to put one's best foot forward, is a contributing factor. The EQ has been found to be the AT-SAT subtest most susceptible to coaching interventions (Heil et al., 2002). In other words, it is relatively easy to teach test takers to present themselves in a positive light on this subtest to help create the illusion that they possess the type of workplace behaviors that are associated with successful ATCSs. Each EQ subscale is comprised of unique items not shared with other EQ subscales. The high intercorrelations among EQ subtests are, therefore, not being driven by shared (overlapping) items. Given these high intercorrelations, it is not surprising that only one factor could be extracted. Previous research using 330 participants (Quartetti et al., 2001) suggested the possibility of a two-factor solution. That version of the EQ, however, contained 201, as opposed to 135, items. Longitudinal research will define the usefulness of the EQ in predicting success as an ATCS. Such research should determine the incremental validity of the EQ over a test of normal personality, which measures more than a single factor. The role of impression management is another area worthy of additional investigation. The EQ subtests appear to lack specificity. For this reason, use of individual EQ subtests would be psychometrically inappropriate.

Table 4. Correlation of NEO PI-R domain scales with Overall EQ score.

	Overall EQ Score
Neuroticism	-.34**
Extraversion	.02
Openness to Experience	.04
Agreeableness	.19*
Conscientiousness	.35**

* $p < .05$ (2-tailed)

** $p < .01$ (2-tailed)

Study 3: Reliability and Validity in Select-In Testing with Research Tests

Most recently, personality research at CAMI has included administration of the (United States Air Force, USAF) Armstrong Laboratory Aviator Personality Survey (ALAPS; Retzlaff, King, Callister, Orme, & Marsh, 2002). The ALAPS was developed through the integration of clinical theory, psychometric methods, and empirical testing. It is composed of 15 scales that assess personality, psychopathology, and crew interaction. It has been demonstrated to be reliable and has been administered to over 6,000 USAF student pilots (Retzlaff et al., 2002), as depicted in Table 5.

The USAF is currently testing the validity of the ALAPS as a selection and cockpit (fighter vs. transport aircraft) assignment tool. Also, Berg, Moore, Retzlaff,

Table 5. USAF Student Pilot ALAPS scales with ranges, means, standard deviations, and Cronbach alphas.

Personality	Range*	Mean	S.D.	Cronbach alpha
Confidence	0 – 16	9.65	2.98	.73
Socialness	0 – 16	12.56	3.41	.83
Aggressiveness	0 – 16	9.29	2.99	.73
Orderliness	0 – 16	11.98	3.53	.84
Negativity	0 – 16	5.51	3.19	.75
Psychopathology				
Affective Lability	0 –16	5.09	3.97	.86
Anxiety	0 –16	2.49	3.49	.90
Depression	0 –16	1.82	2.48	.83
Alcohol Abuse	0 –16	7.65	4.07	.87
Crew Interaction				
Dogmatism	0 – 16	5.83	3.01	.75
Deference	0 – 16	6.31	2.81	.73
Team Oriented	0 – 16	11.90	3.76	.87
Organization	0 – 16	12.40	3.40	.84
Impulsivity	0 – 16	7.36	3.64	.81
Risk Taking	1 – 16	12.21	2.94	.76

*All scales have 16 items.

and King (2002) administered the ALAPS to 312 US Navy aviators and found differences between junior and senior officers.

For a personality measure to be incorporated as part of a screening process, the scores on the measure must be demonstrated to be both psychometrically sound and related to employee performance (Hogan, Hogan, & Roberts, 1996). By including scales gauging crew/team interaction styles, ALAPS extends the traditional limits of most psychological tests. These qualities may be important in the work of ATCSs, if they can be tied back to the traits and job tasks identified by the SMEs as described in Study 2. The purpose of Study 3, therefore, was to provide initial evidence of the psychometric qualities and potential utility of the ALAPS in the ATCS career field.

METHOD

Participants. A total of 121 participants — 102 (84.3%) male and 19 (15.7%) female — who were students at the FAA Academy participated in the study. Their ages ranged from 21 to 34 years (mean age of 27.27 years).

Materials. The ALAPS is a 240-item test with a true/false format, composed of 15 scales (each with 16 items), as outlined in Table 5. The ALAPS includes scales across three major areas: Personality (five scales), Psychopathology (four scales), and Crew/Team Interaction (six scales).

As described in Study 2, participants voluntarily took AT-SAT, which contains the EQ, a subtest comprised of 135 items that assess an individual's work-related attributes based on his/her self-reported past experiences.

Procedure. The participants in this study voluntarily took AT-SAT as part of a research study to longitudinally assess the value of AT-SAT for the selection of ATCSs. They further consented to the ALAPS for the purpose of gauging the potential role of temperament and personality in predicting occupational success. Students were administered the AT-SAT, ALAPS, NEO PI-R, biographical questionnaires, and several other paper-and-pencil tests upon the commencement of their training at the FAA Academy.

RESULTS

The ALAPS data will be looked at from descriptive, reliability, and validity perspectives. As depicted in Table 6, student ATCS ALAPS scores, while similar in overall range to those of USAF student pilots (Table 5), were more constrained on some dimensions. The most constrained variable (*Anxiety*) had an 11-point range in this ATC sample as compared with the full range of 17 obtained from the USAF sample. The student ATCSs exhibited a full range of scores on the *Aggression* scale. Scores on several of the other scales, *Alcohol Abuse, Impulsivity,* and *Risk Taking,* nearly encompassed the full range of 17. As such, few ALAPS scales had a "ceiling" or "floor" effect.

Further, means (Table 6) range from .98 (*Depression*) to 13.0 (*Team Orientation*), with a median of 7.6. These results show a potential to differentiate ATCS applicants.

Table 6. FAA Academy ATCS students' ALAPS scales with ranges, means, standard deviations, and Cronbach alphas.

Personality	Range*	Mean	S.D.	Cronbach alpha
Confidence	2 – 16	9.10	2.66	.67
Socialness	2 – 16	12.98	2.95	.81
Aggressiveness	0 – 16	7.60	2.88	.75
Orderliness	2 – 16	12.11	3.17	.80
Negativity	0 - 12	3.41	2.66	.70
Psychopathology				
Affective Lability	0 –12	3.49	2.71	.72
Anxiety	0 –10	1.12	1.93	.78
Depression	0 –13	.98	1.58	.69
Alcohol Abuse	0 –15	5.82	3.64	.84
Crew Interaction				
Dogmatism	0 – 11	3.88	2.36	.64
Deference	0 – 14	7.88	2.92	.69
Team Oriented	3 – 16	13.00	3.26	.86
Organization	3 – 16	12.82	2.87	.79
Impulsivity	0 – 15	5.38	3.46	.81
Risk Taking	1 – 16	9.12	3.52	.78

* All scales have 16 items.

The means also appear different from those of USAF student pilots. For example, student ATCS scores were significantly lower than those of student pilots on *Aggressiveness* (t=5.77, p<.001), *Impulsivity* (t=5.61, p<.001), and *Risk Taking* (t=8.93, p<.001).

The scales appeared to be reliable with Cronbach alphas ranging from a low of .64 (*Dogmatism*) to a high of .86 (*Team Oriented*). Only two of the 15 scales had reliabilities below .69. Indeed, five of the scales had Cronbach alphas that were above .80. These values were comparable with the reliabilities found in the USAF sample.

To demonstrate the relationship with skills and abilities relevant for ATCSs, the scores obtained on the ALAPS scales were correlated with the overall AT-SAT score. The scale *Depression* correlated with the composite AT-SAT scores at -.26 (p<.01), and *Organization* correlated with AT-SAT at .18 (p<.05). These significant correlations suggest that there are dimensions of ALAPS that are related to the overall skills and abilities necessary for individuals to achieve success as ATCSs. Future studies will examine the relationship of the ALAPS scales with AT-SAT subtest scores, which are currently not accessible because of the recent reweighting of the subtests.

DISCUSSION

The ALAPS scores possess a number of important psychometric characteristics. First, the scales had large ranges and logical means in the samples presented here. There was less variability in the ATCS scores than in the USAF data, perhaps as a function of the restricted nature of the ATCS sample.

Second, the ALAPS scales are, by and large, reliable. They are at least as reliable as the facet scales of the NEO-PI-R, as reported by Costa and McCrae (1992).

Third, ALAPS includes dimensions that may be relevant to performance in the operational ATC environment. It is all too often the case that commercial off-the-shelf tests do not have many of the scales relevant to a dynamic and time-critical environment. It is also doubtful that tests normed for traditional clinical use would also work well with high-functioning ATCSs. Dimensions on the ALAPS (such as aggressiveness, confidence, impulsivity, being team oriented, and a risk taker) are all relevant to aspects of how individuals perform and how well they work closely with others. Further work needs to be done to assess the extent to which the ALAPS scales gauge the qualities that the SMEs deemed to be important in ATCS duties.

Finally, scores on several of the ALAPS scales were related to overall performance on the test (AT-SAT) currently in use to select future ATCSs. The correlation between scores on the ALAPS scale *Depression* and the AT-SAT composite score suggests that ATCS students who are relatively unhappy and/or pessimistic do less well than their more upbeat colleagues on AT-SAT. Conversely, the correlation between scores on the ALAPS *Organization* scale and the AT-SAT composite score indicates that individuals who report a more structured approach to life perform better on AT-SAT and may have greater potential to do well in training and on the job. Alternatively, perhaps motivation and/or reading ability underlie success on both tests. The recent addition of the ALAPS to FAA research efforts extends the half-century of collaboration between the USAF and the FAA.

CONCLUSION

A historical review affords the reader an appreciation of the uses of personality measures in personnel selection that fell out of use in favor of cognitive tests until the relatively recent development of factorially derived measures of personality (Costa & McCrae, 1992). The FAA is continually evaluating the success of cognitive tests and time-shared tasks, as well as personality measures, as part of selection procedures. The goal is to identify candidates who possess the necessary knowledge, skills, abilities, and temperament to successfully control air traffic and who will continue to function efficiently throughout their careers. Future research endeavors will, as part of the overall effort to conduct a longitudinal validation of AT-SAT, gauge the ability of personality tests to predict training outcomes and job performance.

REFERENCES

Americans With Disabilities Act of 1990, Public Law No. 101-336, § 2, 104 Stat. 328 (1991).

Barrick, M.R. & Mount, M.K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1-26.

Berg, J.S., Moore, J.L., Retzlaff, P.D., & King, R.E. (2002). Assessment of personality and crew interaction skills in successful Naval aviators. *Aviation, Space, and Environmental Medicine, 73*, 575 – 9.

Brokaw, L.D. (1959). School and job validation of selection measures for air traffic control training (WADC-TN-59-39).

Carretta, T.R. & Ree, M.J. (2003). Pilot selection methods. In PS Tsang, MA Vidulich (Eds), *Principles and practice of aviation psychology* (pp. 357 – 96). Mahway, NJ: Erlbaum.

Cattell, R.B., Eber, H.W., & Tatsuoka, M.M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16 PF)*. Champaign, IL: Institute for Personality and Ability Testing.

Cobb, B.B. (1962). Problems in Air Traffic Management: II Prediction of success in Air Traffic Control School (62-2). Washington, DC: FAA Office of Aerospace Medicine. [5]

Collins, W.E., Schroeder, D.J., & Nye, L.G. (1989*). Relationship of anxiety scores to Academy and field training performance of air traffic control specialists* (DOT/FAA/AM-89/7). Washington, DC: FAA Office of Aerospace Medicine. [5]

Convey, J.C. (1984). Personality assessment of ATC applicants. In SB Sells, JT Daily, & EW Pickerel (Eds.), *Selection of Air Traffic Controllers,* pp. 323-51, (FAA-AM-84-2), Washington, DC: U.S. Government Printing Office.

Costa, P.T. & McCrae, R.R. (1992). Professional manual, Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI). Odessa, FL: Psychological Assessment Resources.

Dean, M.A., Russell, C.J., & Farmer, W.L. (2002). Non-cognitive predictors of air traffic controller performance. In H Eibfeldt, MC Heil, & D Broach (Eds), *Staffing the ATM system* (pp. 59-72). Burlington, VT: Ashgate.

Department of Transportation (1969). *Secretary's Annual Report to the President*. Washington, DC.

FAA will require exams for air traffic personnel. (1966, February 1), *U.S. Medicine*, 2, pp. 1, 17.

Federal Aviation Administration (1965, October 15). FAA Order 9430.2, *Establishment of Health Program for Agency Air Traffic Control Specialist Employees*. Washington, DC: U.S. Government Printing Office.

Guion, R.M. & Gottier, R.F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, *18*, 135-64.

Heil, M.C., Detwiler, C.A., Agen, R., Williams, C.A., Agnew, B., & King, R.E. (2002). *The effects of practice and coaching on the Air Traffic Selection and Training test battery training* (DOT/FAA/AM-02/24). Washington, DC: FAA Office of Aerospace Medicine. [5]

Hogan, R., Hogan, J., & Roberts, B.W. (1996). Personality measurement and employment decisions. *American Psychologist, 51*, 469 - 77.

Hough, L.M. (1992). The "big five" personality variables – construct confusion: Description versus prediction. *Human Performance, 5*, 139-55.

Kraft, M.A. (1958). *Physical qualifications of air traffic control personnel*. New York: Flight Safety Foundation.

Mischel, W. (1969). Continuity and change in personality. *American Psychologist*, *24*, 1012 – 18.

Nickels, B.J., Bobko, P., Blair, M.D., Sands, W.A., & Tartak, E.L. (1995). *Separation and control assessment (SACHA) final job analysis report*. Bethesda, MD: University Research Corporation.

Paullin, C., Houston, J., Bruskiewicz, K., & McKee, A. (1992). *Draft report on noncognitive inventory development*. Minneapolis, MN: Personnel Decisions Research Institutes, Inc.

Quartetti, D., Waugh, G., Graves, J.G., Abrahams, N.M., Kieckhaefer, W., Houston, J., & Wise, L. (2001). Analysis and revisions of the AT-SAT pilot test. In RA Ramos, MC Heil, & CA Manning (Eds), *Documentation of validity for the AT-SAT computerized test battery, Volume 1* (DOT/FAA/AM-01/5). Washington, DC: FAA Office of Aerospace Medicine, 37-59. [5]

Retzlaff, P.D., King, R.E., Callister, J.D., Orme, D.R., & Marsh, R.W. (2002). The Armstrong Laboratory Aviation Personality Survey: Development, norming, and validation. *Military Medicine*, *167*, 1026 - 32.

Ryan, A.M., Ployhart, R.E., & Friedel, L.A. (1998). Using personality testing to reduce adverse impact: A cautionary note. *Journal of Applied Psychology, 83*(2), 298 - 307.

[5] This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications Web site: http://www.cami.jccbi.gov/aam-400A/index.html

Schmidt, F.L., Ones, D.S., & Hunter, J.E. (1992). Personnel selection. *Annual Review of Psychology*, *43*, 627-70.

Schneider, B., Smith, D.B., Taylor, S., & Fleenor, J. (1998). Personality and organizations: A test of the homogeneity of personality hypothesis. *Journal of Applied Psychology, 83,* 462-70.

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: FAA Office of Aerospace Medicine. [5]

Schwarzkopf, A., Buckley, M.R., & Pace, T.M. (1996). *16PF screen evaluation* (DTFA-02-93-D-93088). Unpublished manuscript, University of Oklahoma.

Selection and Classification Branch (1958). *On-the-job validation of certain measures for selection of air traffic control trainees* (WCLL-TM-58-4). Washington, DC: U.S. Government Printing Office.

Senate Report No. 56-310 (1965).

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

Trites, D.K. (1961). Problems in air traffic management: I. Longitudinal prediction of effectiveness of air traffic controllers (61-1). Washington, DC: FAA Office of Aerospace Medicine.