Federal Aviation Administration

DOT/FAA/AM-13/3
Office of Aerospace Medicine
Washington, DC 20591

The Validity of the Air Traffic Selection and Training (AT-SAT) Test Battery in Operational Use

Dana Broach
Cristina L. Byrne
Carol A. Manning
Linda Pierce
Darendia McCauley
M. Kathryn Bleckley

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

March 2013

Final Report

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-13/3		

4. Title and Subtitle	5. Report Date
The Validity of the Air Traffic Selection and Training (AT-SAT) Test Battery in Operational Use	March 2013
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Broach D, Byrne CL, Manning CA, Pierce L, McCauley D, Bleckley MK	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
FAA Civil Aerospace Medical 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 Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	14. Sponsoring Agency Code

15. Supplemental Notes

16. Abstract

Applicants for the air traffic control specialist (ATCS) occupation from the general public and graduates from post-secondary institutions participating in the FAA's Air Traffic Collegiate Training Initiative (AT-CTI) must take and pass the Air Traffic Selection and Training (AT-SAT) test battery as part of the selection process. Two concurrent, criterion-related validation studies demonstrated that AT-SAT was a valid predictor of ATCS job performance (American Institutes for Research, 2012; Ramos, Heil, & Manning, 2001a,b). However, the validity of AT-SAT in operational use has been questioned since implementation in 2002 (Barr, Brady, Koleszar, New, & Pounds, 2011; Department of Transportation Office of the Inspector General, 2010). The current study investigated the validity of AT-SAT in operational use.

Method. AT-SAT and field training data for 1,950 air traffic controllers hired in fiscal years 2007 through 2009 were analyzed by correlation, cross-tabulation, and logistic regression with achievement of Certified Professional Controller (CPC) status as the criterion.

Results. The correlation between AT-SAT and achievement of CPC status was .127 (n=1,950, p<.001). The correlation was .188 when corrected for direct restriction in range. A larger proportion of controllers in the "Well Qualified" score band (85-100) achieved CPC status than in the "Qualified" (70-84.99) band. The logistic regression model did not fit the data well (χ^2=30.659, p<.001, -2LL=1920.911). AT-SAT modeled only a small proportion of the variance in achievement of CPC status (Cox and Snell R^2=.016, Nagelkerke R^2=.025). The logistic regression coefficient for AT-SAT score of .049 was significant (Wald=30.958, p<.001).

Discussion. AT-SAT is a valid predictor of achievement of CPC status at the first assigned field facility. However, the correlation is likely attenuated by time and intervening variables such as the training process itself. Other factors might include the weighting of subtest scores and use of a narrow criterion measure. Further research on the validity of AT-SAT in relation to multiple criteria is recommended.

17. Key Words	18. Distribution Statement
ATCS Selection, Aptitude,Validity	Document is available to the public through the Internet: www.faa.gov/go/oamtechreports

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	14	

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

ACKNOWLEDGMENTS

Research reported in this paper was conducted under the Air Traffic Program Directive/ Level of Effort Agreement between the Human Factors Research and Engineering Division (ANG-C1), FAA Headquarters, and the Aerospace Human Factors Research Division (AAM-500) of the FAA Civil Aerospace Medical Institute.

The opinions expressed are those of the authors alone, and do not necessarily reflect those of the Federal Aviation Administration, the Department of Transportation, or Federal government of the United States.

Correspondence concerning this report should be addressed to Dana Broach, Aerospace Human Factors Research Division (AAM-500), P.O. Box 25082, Oklahoma City, OK 73125. E-mail: dana.broach@faa.gov

CONTENTS

The Validity of the Air Traffic Selection and Training (AT-SAT) Test Battery in Operational Use-------- 1

Background ---------- 2

Method ---------- 2

Sample ---------- 2

Measures---------- 3

Analyses---------- 4

Results---------- 5

Discussion ---------- 6

References---------- 8

The Validity of the Air Traffic Selection and Training (AT-SAT) Test Battery in Operational Use

The air traffic control specialist (ATCS) occupation is the single largest and most publicly visible workforce in the Federal Aviation Administration (FAA). ATCSs, also known as air traffic controllers, or most simply, controllers, are responsible for the safe, efficient, and orderly flow of air traffic in the U.S. air transportation system. There are just over 15,000 non-supervisory controllers working in 315 air traffic control facilities handling 30,000 commercial and other flights per day. It is an attractive job with a six-figure income and federal benefits – if a person survives the winnowing process from application to certification. In the past, less than 4% of applicants successfully completed the grueling gauntlet of aptitude tests, screens, classroom training, simulation training, on-the-job training, and over-the-shoulder performance evaluations with live traffic to become fully certified controllers (Broach, 1998).

The first hurdle in this lengthy process is getting hired. The FAA projects that it will hire several hundred to about a thousand new controllers each year between now and 2020 to replace retiring controllers (FAA, 2012). There are three primary paths to becoming an air traffic controller with the FAA. The first path is for persons who have previously been appointed and served as controllers. According to ATCS hiring data compiled by the Air Traffic Organization (R. Mitchell, personal communication, October 17, 2012), about 30% of new controllers have entered the FAA via this path in recent years, most commonly from the ranks of military air traffic controllers. The second path is for persons who completed an ATCS training program at one of 36 post-secondary educational institutions participating in the FAA's Air Traffic Collegiate Training Initiative (AT-CTI) program. About 35% of new controllers entered the FAA via the AT-CTI path since 2006. The third path is for persons from the general public. No previous air traffic control experience or training is required on this path. About 35% of new controllers have entered service via the general public path since 2006. There are several other paths, but they account for a very small proportion of new hires since 2006. The focus of this paper is on those hired via the AT-CTI and general public paths.

The U.S. Office of Personnel Management (OPM) established the qualification standards that an applicant must meet to enter the ATCS occupation. At a bare minimum, an ATCS applicant must be a U.S. citizen and have a high school diploma (or equivalent). In addition, the applicant must have three years of progressively responsible work experience or a four-year college degree or some combination of work experience and post-secondary education. A general public or AT-CTI applicant must also meet two additional qualification standards. First, an applicant must not have reached his or her 31st birthday by the time a bona fide tentative job offer is made and accepted. Second, the applicant must obtain a qualifying score on an aptitude test for the occupation.

The computerized Air Traffic Selection and Training (AT-SAT) battery is the aptitude test currently used by the FAA to assess general public and AT-CTI applicants under the OPM occupational qualification standard. AT-SAT has been in operational use since 2002 (King, Manning, & Drechsler, 2007). Relatively few persons were tested in 2002 through 2005, as the FAA was not hiring many new air traffic controllers at that time. However, beginning in mid-2006, retirements from the ATCS workforce surged, and the pace of hiring new controllers increased substantially. Since 2006, FAA has administered the AT-SAT battery to more than 22,000 applicants and hired 6,826 as new controllers via the AT-CTI and general public paths.

Three principal criticisms of AT-SAT have been made. First, significant differences in score distributions by race and sex were observed in the course of validation, with Blacks and Hispanic-Latinos scoring lower than Whites and women scoring lower than men (Waugh, 2001, p. 44). The FAA re-weighted the AT-SAT subtests to mitigate these group differences without substantially reducing validity (Wise, Tsacoumis, Waugh, Putka, & Hom, 2001; Dattel & King, 2006; King, Manning, & Drechsler, 2006). Second, the pass rate was substantially higher than was originally projected. While a pass rate of about 67% was predicted by Wise et al. after re-weighting, the actual pass rate in operational use has been greater than 90% (Department of Transportation Office of the Inspector General [DOT OIG], 2010; King et al.). Third, the validity of AT-SAT as a predictor of training outcomes and job performance has been questioned. For example, a Congressional committee chairman has expressed particular concern "about whether FAA's screening test identifies candidates' potential to become air traffic controllers" (DOT OIG, 2010, p. 1). Barr, Brady, Koleszar, New, & Pounds (2011) found no "completed studies that determined if the AT-SAT actually predicted job performance success among those who took the exam, were accepted for Academy training, and who subsequently entered and completed on-the-job training in the field" (p. 9). They concluded that without such a longitudinal study, "the FAA cannot be sure that the

AT-SAT is accomplishing its original goals of predictability" (ibid). The purpose of the current study is to investigate the validity of AT-SAT as a predictor of training outcomes: To what degree does AT-SAT predict successful completion of on-the-job training in the field?

BACKGROUND

"Validity" is used here in accordance with the Uniform Guidelines on Employee Selection Procedures (29 C.F.R. § 1607 (2012)), the Civil Rights Act of 1991 (42 U.S.C. § 2000e et seq., 2011), and the relevant professional standards and principles for the development, validation, and use of employee selection test and procedures (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Society for Industrial and Organizational Psychology (SIOP), 2003). Validity refers to the evidence supporting the inference to be drawn on the basis of a score on a given test. In personnel selection, the inference to be drawn is expected future job performance, as represented by criterion measures (Sackett, Putka, & McCloy, 2012; SIOP, 2003). Example criterion measures representing job performance are production rate, error rate, tenure (retention), job performance ratings, and training performance, including outcomes (14 C.F.R. § 1607.14B(3)). That predictive inference about future job performance is made on the basis of the statistical relationship between predictor test score and the criterion measure, where the relationship is expressed as a correlation. Also known as a validity coefficient, the correlation mathematically describes how much the criterion measure changes as a function of predictor test scores.

Validation is the process of accumulating empirical evidence about that statistical relationship between test score and the criterion (or criteria) to support the predictive inference. Two common approaches to developing validation evidence in personnel selection contexts are predictive criterion-related validation studies and concurrent criterion-related validation studies (SIOP, 2003). Empirical evidence for the validity of AT-SAT as a predictor of job performance was provided through two concurrent, criterion-related validation studies. The first study was reported in 2001 (Ramos, Heil, & Manning, 2001a, b). Approximately 1,000 incumbent en route controllers took the proposed test battery. Job performance data were collected concurrently in two forms: Behavioral Summary Scale (BSS) ratings of job performance by peers and supervisors; and the en route Computer-Based Performance Measure (CBPM; see Hanson, Borman, Mogilka, Manning, & Hedge, 1999). The correlation between the test score and the composite job performance measure was .52 without any corrections for range restriction or criterion unreliability. With correction for incidental range restriction, the correlation was .68 (Waugh, 2001). The second concurrent criterion-related validation study was conducted by the American Institutes for Research (AIR; 2012). The current operational version of AT-SAT was administered to 302 incumbent air traffic control tower (ATCT) controllers. As in the original en route validation study, two classes of job performance data were collected: Behavioral Summary Scale (BSS) ratings of job performance by peers and supervisors; and performance on the Tower Computer-Based Performance Measure (see Horgen, et al., 2012). The correlation between an optimally-weighted composite of AT-SAT subtest scores and the composite of the two criterion measures was .42 without any corrections (AIR, p. 47). These two studies independently demonstrated that AT-SAT is a valid predictor of ATCS job performance. The current study develops a third line of evidence for the validity of AT-SAT by investigating the degree to which achievement of CPC status at the first field facility can be predicted from AT-SAT scores.

METHOD

Sample

The sample for this study consisted of air traffic controllers hired in fiscal years 2007-2009. Sufficient time has elapsed for most persons hired in these fiscal years to complete the field training sequence, averaging two to three years. To identify the sample, records were extracted from the Air Traffic Organization's Air Traffic Controller National Training Database (ATC NTD) and matched with AT-SAT examination records at the individual level. The ATC NTD contains data for persons who reported to a field facility for on-the-job training (OJT); data for persons who failed or withdrew from FAA Academy training and did not enter OJT at a field facility are not in the NTD. The ATC NTD contained records for 11,450 new hires at field facilities as of July 2012, of which 6,941 were for general public or CTI hires. This pool was reduced to 6,865 records after screening for complete identifiers and duplicates. These records were then filtered by fiscal year of entry-on-duty and valid AT-SAT scores, resulting in a sample of 2,569 first facility training records for new controllers. Records for new hires who left the field facility training for other reasons (unrelated to performance, per NTD; n=160), who requested transfer prior to completion of facility training (n=156), or who were still in facility training (n=303) were dropped, leaving a total of 1,950 records for analysis.

All of the controllers in the sample had been hired under vacancy announcements open to the general public and AT-CTI graduates. Most (69%) were hired under a general public announcement. The sample was predominantly

Table 1. Demographic characteristics and descriptive statistics

Characteristic	Applicants (*N*=15,173)	Sample (*N*=1,950)
Race/National Origin (RNO) Group		
Asian	464 (3.1%)	45 (2.3%)
Black	3,039 (20.0%)	175 (9.0%)
Hawaiian-Pacific Island	77 (0.5%)	6 (0.3%)
Hispanic-Latino	814 (5.4%)	65 (3.3%)
Native American-Alaskan Native	63 (0.4%)	10 (0.5%)
White	8,906 (58.7%)	1,1,73 (60.2%)
Multi-racial[1]	1,059 (7.0%)	102 (5.2%)
No RNO group(s) marked	738 (4.9%)	96 (4.9%)
Missing data	13 (0.1%)	278 (14.3%)
Sex		
Female	3,449 (22.7%)	307 (15.7%)
Male	11,127 (73.3%)	1,330 (68.2%)
Missing data	597 (3.9%)	313 (16.1%)
Age Mean (SD)	25.2 (3.25)	25.2 (2.84)
AT-SAT Mean (SD)	85.87 (9.39)	90.99 (6.27)

Notes: [1]Two or more RNO groups marked

Table 2. AT-SAT Subtests

Subtest	Description
Dials (DI)	Scan and interpret readings from a cluster of analog instruments
Applied Math (AM)	Solve basic distance, rate, and time problems
Scan (SC)	Scan dynamic display to detect targets that change over time
Angles (AN)	Determine interior and exterior angle of intersecting lines
Letter Factory (LF)	Manage virtual production line, box products, perform quality control
Air Traffic Scenarios Test (ATST)	Direct aircraft to destination in low-fidelity radar simulation
Analogies (AY)	Solve verbal and non-verbal analogies
Experience Questionnaire (EQ)	Life experiences, preferences, and typical behavior in situations

White (60%) and male (68%). The average age at the time of entry-on-duty with the FAA was 25.2 (SD=2.8 years). Demographic statistics for the general public and CTI applicant population (n=15,173) and the sample are presented in Table 1.

Measures

AT-SAT is a computerized aptitude test of cognitive abilities, skills, and other personal characteristics identified through formal job analysis as being required at the time of entry into the ATCS occupation. AT-SAT consists of eight subtests: Dials (DI); Applied Math (AM); Scan (SC); Angles (AN); Letter Factory Test (LF); Air Traffic Scenarios Test (ATST); Analogies (AY); and the Experience Questionnaire (EQ). See Table 2 for a brief description of each subtest. A weighted composite score is computed from subtest scores.

The FAA uses category ranking in the selection of controllers. Applicants with AT-SAT composite scores of less than 70 are not qualified for consideration for employment. Scores of 70 to 84.99 place an applicant in the "Qualified" category, while scores of 85 to 100 put an applicant in the "Well Qualified" category. Applicants in the "Well Qualified" category are considered first for vacancies, with veteran's preference applied in accordance with civil service rules. Applicants in the "Qualified" category are considered if there is an insufficient number in the "Well Qualified" category to meet FAA hiring needs. As a consequence, the sample (persons hired) was largely drawn from the "Well Qualified" category (87% of the sample). However, among all applicants, just 57% are ranked in the "Well Qualified" category. In other words, "Well Qualified" candidates were over-represented in the sample relative to the applicant population. The mean AT-SAT score for the sample was 86.29 (SD=6.42), compared to 85.85 (SD=9.41) for the applicant population.

Table 3. Training outcome at first field facility as coded in the NTD

	N	%
Not CPC: Facility Fail		
Employment Terminated Prior to Completion	135	6.0%
Reassigned to a non-ATC FAA position	22	1.0%
Training Discontinued by Air Traffic Manager (ATM)	10	0.4%
Training Failure - Pending Human Resources (HR) Action	5	0.2%
Employment Termination Letter Issued	4	0.2%
Employee Withdrew From Training	2	0.1%
Not CPC: Transfer Lower		
Reassigned to Another 2152 Facility	212	9.4%
CPC: Successfully Completed Training	1,560	69.2%

Successful completion of training at the first field facility is a desirable outcome for both the agency and the individual. Therefore, a dichotomous variable was derived from the ATC NTD data to represent success in first facility OJT. Individuals who failed, withdrew, or had training terminated at the first facility were coded as "Facility Fail" in the ATC NTD. Facility failure can result in the termination of employment. However, the agency also has the discretion to offer an individual at risk for failure a transfer to a lower level, less complex facility if a position is available (FAA, 2006). The ATC NTD coded these cases as "Transfer Lower." Such a transfer is still an adverse outcome from an agency perspective due to the associated costs and staffing gap created by the loss at the first facility. Therefore, persons who were categorized in NTD as "Facility Fail" and "Transfer Lower" were coded as having failed to achieve CPC status ("Not CPC"; n=390) at the first facility, while persons categorized as "Completed" in the ATC NTD (n=1,560) were coded as "CPC" (Table 3).

Analyses

Three analyses were conducted. First, the simple Pearson product-moment correlation between AT-SAT score and field training outcome (achievement of CPC status) at the first assigned field facility was computed, without corrections for direct range restriction on the predictor or criterion unreliability. This raw correlation provides a conservative, lower-bound estimate of AT-SAT's validity as a predictor of field training outcome. The correlation was then corrected for direct range restriction on the predictor (AT-SAT) using the Ghiselli, Campbell, and Zedeck (1981, p. 299) equation 10-12. The corrected correlation provides a less biased estimate of the AT-SAT's validity as a predictor of field training outcome. No correction for criterion unreliability was made. Second, a 2-by-2 (AT-SAT score band [Qualified, Well Qualified] by first facility training outcome [Not CPC, CPC]) $\chi 2$ analysis was conducted. The odds of certifying by score band were estimated. Third, logistic regression was used to model the relationship of AT-SAT score to achievement of CPC status at the first field facility. The odds of certifying by AT-SAT score were estimated from the logistic regression equation. All analyses were conducted using SPSS version 20.

RESULTS

The simple correlation between AT-SAT score and achievement of CPC status at the first field facility was .127 (n=1,950, p <.001) without corrections for direct range restriction on the predictor or criterion unreliability. With correction for direct range restriction on the predictor, the correlation between AT-SAT score and status was .188. The usual tests of statistical significance do not apply to correlations corrected for restriction in range (SIOP, 2003).

The cross-tabulation of AT-SAT score band by achievement of CPC status is presented in Table 4. Eight persons scored below 70 on their first AT-SAT examination and were excluded from the cross-tabulation analysis. Because of FAA hiring policies, most new hires were selected from the Well Qualified score band. Overall, 82% of the 1,681 Well Qualified new hires successfully completed field training at the first assigned field facility, compared to 71% of the 261 new hires from the Qualified score band (χ^2=17.54, p <.001). New hires from the Well Qualified score band were 1.86 times more likely to achieve CPC status than new hires from the Qualified score band (odds ratio confidence interval=1.39 to 2.49).

The logistic regression of AT-SAT on achievement of CPC status at the first facility resulted in correct classification of 57.5% of the 1,950 cases, as shown in Table 5. As expected, in view of the modest correlation between AT-SAT and the criterion measure, the model did not fit the data well (χ^2=30.66, p <.001, -2LL=1920.91). AT-SAT modeled only a small proportion of the variance in the field training outcome of achieving CPC status (Cox and Snell R^2=.016, Nagelkerke R^2=.025).

Table 4. Cross-tabulation of AT-SAT score band by field training outcome (expectancy table)

	Field Training Outcome		
AT-SAT Score Band	Unsuccessful	Successful	Total
Qualified	77	184	261
	(29.5%)	(70.5%)	
Well Qualified	309	1,372	1,681
	(18.4%)	(81.6%)	
Total	386	1,556	1,942

Table 5. Logistic regression cross-classification table (cut value=.80)

	Predicted Outcome		
Observed Outcome	Unsuccessful	CPC	% Correct
Unsuccessful	217	173	55.6%
CPC	656	904	57.9%
Overall %			57.5%

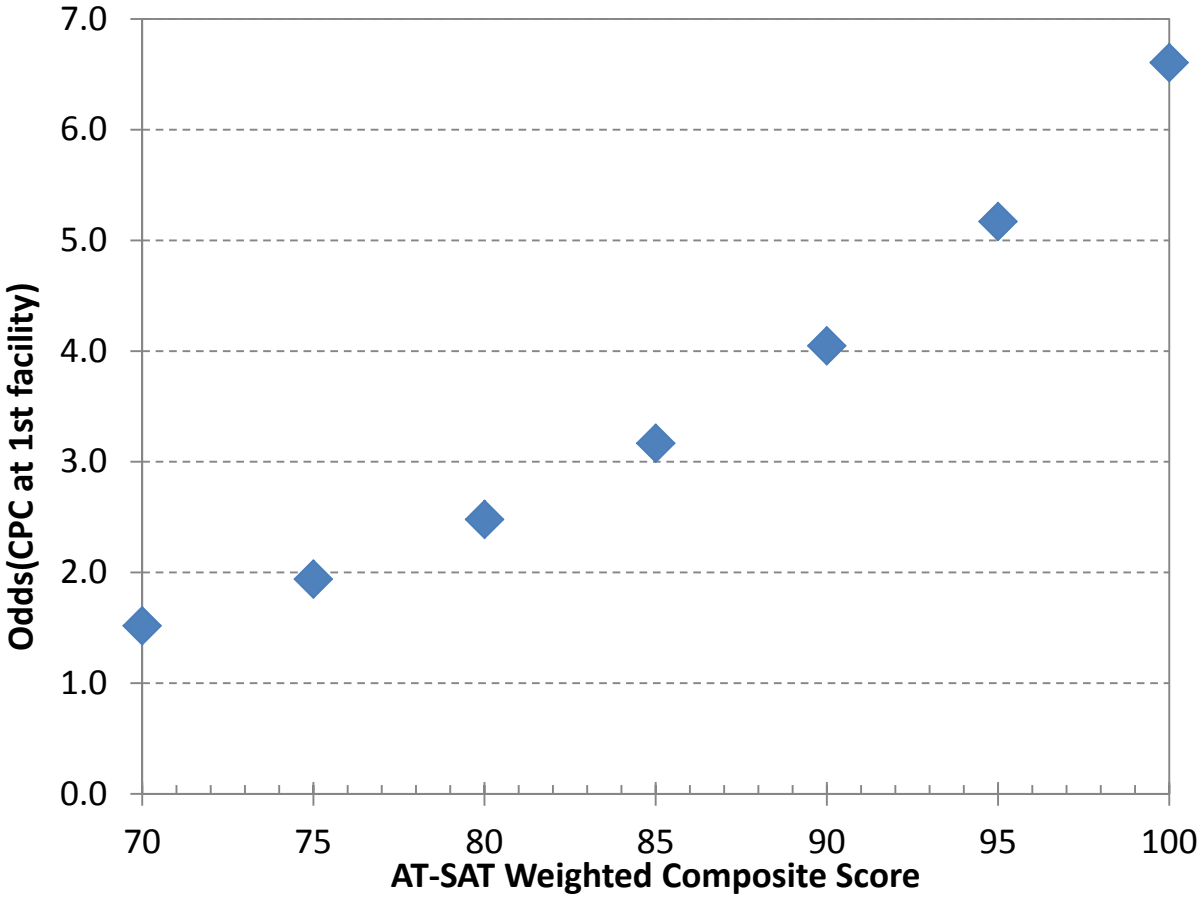

Figure 1. Odds of achieving CPC at the first field facility by AT-SAT composite score

Nevertheless, the logistic regression coefficient for AT-SAT score of .049 was significant (Wald=30.958, p <.001). The odds of certifying at the first assigned field facility were computed from the logistic regression equation as a function of AT-SAT score (Figure 1; see Norusis, 1990, pp.49-50). A new hire with an AT-SAT score of 70 had slightly better than even odds (1.5 to 1) of achieving CPC status. In comparison, a new hire with an AT-SAT score of 85 had slightly better than 3-to-1 odds of achieving CPC status. In other words, new hires with higher AT-SAT scores had better odds of achieving CPC status at the first field facility than new hires with lower AT-SAT scores.

DISCUSSION

The current study investigated the validity of AT-SAT as a predictor of achievement of CPC status at the first field facility. The results showed that AT-SAT was a valid predictor of training outcome for next generation of air traffic controllers. First, the correlation between AT-SAT score and training status was positive and significant. Second, persons with higher scores were more likely to certify at the first assigned field facility than were persons with lower scores as shown by the χ^2 analysis. Third, logistic regression analysis found the odds of certifying at the first facility increased with AT-SAT score. Taken together, the results of the present investigation and those of the two previous criterion-related validation studies show that AT-SAT is a valid predictor of both OJT outcome (achievement of CPC status) and, more importantly, of on-the-job performance after certification. In other words, the empirical evidence supports the validity of AT-SAT as a personnel selection procedure for the ATCS occupation.

The uncorrected correlation between AT-SAT and achievement of CPC status in this study was "small" in Cohen's 1988 frequently cited categorization of effect sizes. In comparison, Bertua, Anderson, and Salgado (2005) reported average uncorrected correlations from .15 to .30 between various types of cognitive ability tests and criterion measures. Other point estimates of the validity of cognitive ability tests range from .29 to .51 (Bobko, Roth, & Potosky, 1999; Hunter & Hunter, 1984; Schmidt & Hunter, 1998). In another meta-analysis, Robbins, et al. (2004) reported an average correlation of .121 between college admissions test (ACT, SAT) scores and retention in 4-year college programs. While the AT-SAT correlation with field training outcome was low, it is within the range of values reported for other cognitively-loaded selection instruments.

Moreover, AT-SAT predicted achievement of CPC status several years after testing despite many intervening variables. Both time and intervening variables attenuate predictor-criterion relationships (Barrett, Caldwell, & Alexander, 1989; Barrett, Alexander, & Doverspike, 1992; Beier & Ackerman, 2012; Murphy 1989; Van Iddekinge & Ployhart, 2008). The average time between testing and completion of field training or loss was 34 months (*SD*=10.9 months). It might also be the case that not all of the field attrition was due to lack of aptitude. For example, losses might be due to economic factors such as a lack of affordable housing and lifestyle factors (e.g., lengthy commute or the availability of affordable and flexible childcare). Losses for these reasons are unlikely to be predictable from an aptitude test. Better information is needed to understand and categorize losses in field training for future investigations of the validity of AT-SAT.

Even though the correlation was modest and despite the intervening variables, AT-SAT as a selection procedure could have practical utility. ATCS selection is a large-scale, high-stakes selection process. ATCS training is expensive, with an estimated cost per developmental of $93,000 per year (FAA, 2012). Selection of only applicants from the "Well Qualified" score band would have increased the net success rate to 82%, avoiding 77 unnecessary field failures in this cohort. Reducing the field failures by 77 persons would have avoided about $7M ($93,000 x 77 persons) in cumulative lost costs in personnel compensation and benefits for this sample of new hires.[1]

In closing, the current study provides additional empirical evidence that AT-SAT is a valid selection procedure for the ATCS occupation. Persons with higher scores on AT-SAT were more likely to successfully certify at their first field facility. Field attrition among developmental controllers has often been framed as a problem in initial selection and placement. However, only a small proportion of the variance in achievement of CPC status was explained by aptitude test scores collected two or three years earlier, as evidenced by the "small" correlation between AT-SAT and CPC status. There are several possible explanations for this observation. First, achievement of CPC status is a binary criterion representing minimally acceptable performance at the completion of training. Binary criteria inherently limit the value of any correlation as the distribution shifts away from a 50/50 split (Ghiselli et al., 1981). In contrast, multiple criterion measures were used in the concurrent, criterion-related validation studies, measures that encompassed the broad range of controller work behaviors. Those criterion measures assessed typical job performance on multiple dimensions from peer and supervisor perspectives and maximal technical job performance on meaningful interval scales. Further investigation of AT-SAT's validity in relation to additional criterion measures such as performance in FAA Academy initial qualifications training, organizational citizenship behavior, counter-productive work behavior, job knowledge, and post-CPC technical job performance are recommended. This will require the development and collection of psychometrically sound measures of individual controller job performance. Second, the weights given to the subtest scores might not be optimal for predicting achievement of CPC status. AT-SAT was originally weighted to select those whose job performance would be higher than average; a different weighting approach might be required to predict CPC status, a far different criterion. Finer-grained analyses of subtest scores and their weights are recommended in continuing evaluations. Third, the relationship of predictor and achievement of CPC status might be attenuated by time and intervening variables. Research on the training process itself, as delivered at field facilities, and investigations into the reasons developmental controllers do not achieve CPC status are recommended. Careful attention must be given to the reasons why and when new controllers leave field training in order to understand what can be predicted from performance on an aptitude test battery and what cannot.

[1]The actual avoided costs depend on when each ind ividual left field training. The FAA estimates the cost of training at $93,000/year, or $7,750/month. If 47 developmental controllers left training after 10 months, 25 at 20 months, and 5 at 30 months, the avoided lost costs would be (47 x 10 x $7,750) + (25 x 20 x $7,750) + (5 x 30 x $7,750), or $8,680,000. The $7M figure is a rough-order-of magnitude or benchmark estimate based on the assumption that attrition occurs in the first year.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing* (4th Ed.). Washington, DC: American Psychological Association.

American Institutes for Research. (2012, September). *Validate AT-SAT as a placement tool.* (Draft report prepared under FAA contract DTFWA-09-A-80027 Appendix C). Oklahoma City, OK: Federal Aviation Administration Aerospace Human Factors Research Division (AAM-500).

Barrett, G.V., Caldwell, M.S., & Alexander, R.A. (1989). The predictive stability of ability requirements for task performance: A critical reanalysis. *Human Performance*, *2*, 167-181.

Barrett, G.V., Alexander, R.A., & Doverspike, D. (1992). The implications for personnel selection of apparent declines in predictive validities over time: A critique of Hulin, Henry, and Noon. *Personnel Psychology*, *45*, 601-617.

Barr, M., Brady, T., Koleszar, G., New, M., & Pounds, J. (September 22, 2011). *FAA Independent Review Panel on the selection, assignment and training of air traffic control specialists.* Washington, DC: Federal Aviation Administration. http://www.faa.gov/news/updates/media/IRP%20Report%20on%20Selection%20Assignment%20Training%20of%20ATCS%20FINAL%2020110922.pdf

Beier, M.E. & Ackerman, P.L. (2012). Time in personnel selection. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 721-739). New York, NY: Oxford University Press.

Bertua, C., Anderson, N., & Salgado, J.F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology*, *78*, 387–409.

Bobko, P., Roth, P., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors and job performance. *Personnel Psychology*, *52*, 561-589.

Broach, D. (Ed.) (1998). *Recovery of the FAA air traffic control specialist workforce.* (Report No. DOT/FAA/AM-98/23). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd Ed.). Hillsdale, NJ: Erlbaum.

Dattel, A.R. & King, R.E. (2006). *Reweighting AT-SAT to mitigate group score differences.* (Report No. DOT/FAA/AM-06/16). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine.

Department of Transportation Office of the Inspector General. (2010). *Review of screening, placement, and initial training of newly hired air traffic controllers.* (Report No. AV-2010-049). Washington, DC: Author.

Federal Aviation Administration. (2006). *Employment policy for air traffic control specialist in training.* (Human Resources Policy Manual (HRPM) Volume 1: Employment HRPM Supplement EMP-1.14 (ATCS Employment Policy)). Last retrieved November 8, 2012 from the FAA employee website https://employees.faa.gov/org/staffoffices/ahr/program_policies/policy_guidance/hr_policies/hrpm/emp/emp-1-14_sup/

Federal Aviation Administration. (2012, April). *A plan for the future: 10-year strategy for the air traffic control workforce 2012-2021.* Washington, DC: Author.

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

Hanson, M.A., Borman, W.C., Mogilka, H.J., Manning, C., & Hedge, J.W. (1999). Computerized assessment of skill for a highly technical job. In F. Drasgow & J.B. Olson-Buchanan (Eds.), *Innovations in computerized assessment* (pp. 197-220). Mahwah, NJ: Lawrence Erlbaum Associates.

Horgen, K., Lentz, E.M., Borman, W.C., Lowe, S.E., Starkey, P.A., & Crutchfield, J.M. (2012, April). *Applications of simulation technology for a highly skilled job.* Paper presented at the 27th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Hunter, J. & Hunter, R. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72-98.

King, R.E., Manning, C.A., & Drechsler, G.K. (2007). *Operational use of the Air Traffic Selection and Training Battery.* (Report No. DOT/FAA/AM-07/14). Washington, DC: Federal Aviation Administration Office of Aerospace Medicine.

Murphy, K.R. (1989). Is the relationship between cognitive ability and job performance stable over time? *Human Performance*, *2*, 183-200

Norusis, M. J. (1990). *SPSS advanced statistics user's guide.* Chicago, IL: SPSS Inc.

Ramos, R.A., Heil, M.C., & Manning, C.A. (Eds.). (2001a). *Documentation of validity for the AT-SAT computerized test battery, Volume I.* (Report No. DOT/FAA/AM-01/5). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Ramos, R.A., Heil, M.C., & Manning, C.A. (Eds.). (2001b). *Documentation of validity for the AT-SAT computerized test battery, Volume II.* (Report No. DOT/FAA/AM-01/6). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Robbins, S.B., Lauver, K., Huy, L., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261-288.

Sackett, P.R., Putka, D.J., & McCloy, R.A. (2012). The concept of validity and the process of validation. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 91-118). New York, NY: Oxford University Press.

Schmidt, F.L. & Hunter, J.E. (1998). The validity of selection methods in personnel psychology: Practical and theoretical implications for 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of employee selection procedures* (4th Ed.). Bowling Green, OH: Author.

Van Iddekinge, C.H. & Ployhart, R.E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology*, *61*, 871-925.

Waugh, G. (2001). Analysis of group differences and fairness. In R.A. Ramos, M.C. Heil, & C.A. Manning (Eds.), *Documentation of validity for the AT-SAT computerized test battery Volume II* (pp. 43-47). (Report No. DOT/FAA/AM-01/6). Washington, DC: Federal Aviation Administration Office of Aviation Medicine.

Wise, L.L., Tsacoumis, S.T., Waugh, G.W., Putka, D.J., & Hom, I. (2001, December). *Revision of the AT-SAT.* (Report No. DTR-01-58). Alexandria, VA: Human Resources Research Organization (HumRRO).