RELATIONSHIP BETWEEN AIR TRAFFIC SELECTION AND TRAINING (AT-SAT) BATTERY TEST SCORES AND COMPOSITE SCORES IN THE INITIAL EN ROUTE AIR TRAFFIC CONTROL QUALIFICATION TRAINING COURSE AT THE FEDERAL AVIATION ADMINISTRATION (FAA) ACADEMY

By

RONALD SCOTT KELLEY

Bachelor of Science in Business Administration
East Central Oklahoma State University
Ada, OK
1987

Master of Science in Aerospace Administration
Southeastern Oklahoma State University
Durant, OK
2007

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF EDUCATION
May, 2012

RELATIONSHIP BETWEEN AIR TRAFFIC SELECTION AND

TRAINING (AT-SAT) BATTERY TEST SCORES AND

COMPOSITE SCORES IN THE INITIAL EN ROUTE

AIR TRAFFIC CONTROL QUALIFICATION

TRAINING COURSE AT THE FEDERAL

AVIATION ADMINISTRATION

(FAA) ACADEMY

Dissertation Approved:

Dr. Todd Hubbard

Dissertation Advisor

Dr. Fredrick Hanson

Dr. Chad Depperschmidt

Dr. James Key

Outside Committee Member

Dr. Sheryl A. Tucker

Dean of the Graduate College

.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Air Traffic Control Specialists are responsible for the separation of aircraft operating in the National Airspace System, orderly flow of aircraft, and support of national security and homeland defense (Federal Aviation Administration, 2010, p. 2-1-1). These air traffic controllers oversee more than 55,000 commercial, military, and civilian flights each day from a network of air traffic facilities. Over 24 million square miles of airspace is monitored continuously from the Air Traffic Control System Command Center (ATCSCC), Air Route Traffic Control Centers (ARTCCs), Terminal Radar Approach Controls (TRACONs), and Tower Cabs (Booz Allen Hamilton, 2007, p. 3). The U.S. General Accounting Office (1997) has estimated that over 500 million passengers are served each year by the 387 FAA air traffic control facilities (p. 2).

Air traffic controller responsibilities vary by the type of facility they are assigned. The three main types of air traffic facilities are tower, terminal radar, and en route. Air traffic controllers assigned to control towers are responsible for ensuring the separation of aircraft on the airport grounds and in the air, typically within a five mile radius. These controllers manage the flow of traffic during the takeoff and landing phase of flight. Air traffic controllers assigned to terminal radar facilities, or TRACONs, typically control the airspace within a 30 mile radius of an airport. These controllers will direct departing aircraft to their initial routings to their destination airport or sequence arriving aircraft to their final approach routing (U.S. GAO, 1997, pp. 12-13). En route air traffic controllers manage the flow of traffic over large distances between an aircraft's departure airport and destination airport. Routes and altitudes are assigned to the aircraft and coordinated with other controllers along the aircraft's route of flight. These controllers work in en route centers that normally overlie several states and encompass over 100,000 square miles of airspace (U.S. GAO, 1997, pp. 12-13). The en route air traffic controller will be the focus of this paper.

The selection and training of the air traffic controller workforce is a critical element of ensuring the safety and efficiency of the National Airspace System. Currently, there are more than 15,000 air traffic controllers employed by the FAA. In the next ten years, over 73 percent of the workforce will be eligible for retirement (Dunleavy, et al., 2006, p.1). It is critical for the FAA to select and train these new employees as efficiently as possible to keep up with the increased demand on the national airspace system and the growing number of retiring air traffic controllers (Booz Allen Hamilton, 2007, p. 3). Broach and Manning (1997) indicate how important the process can be as follows, "Choosing the wrong person for a job can have visibly disastrous results. Nowhere is this more apparent than in air traffic control, where the consequences of errors may be immediate and catastrophic" (p. 1).

Problem Statement

The Air Traffic Selection and Training (AT-SAT) test battery is a computer delivered pre-employment test given by the FAA to screen potential air traffic control specialist candidates. If selected, these candidates are required to attend the Initial En Route Qualification training course delivered at the FAA Academy. This course teaches and evaluates entry level air traffic control job functions.

A quantitative relationship between the AT-SAT test battery and the Initial En Route Qualification training course has not been possible due to the lack of quantitative course scores. Following the development of the redesigned Initial En Route Qualification training course, it is now possible to determine the relationship between the composite score and the individual sub-test scores from the AT-SAT test battery to the student's composite score from the newly redesigned Initial En Route Qualification training course.

Purpose Statement

The purpose of this study is to determine the relationship between the AT-SAT test battery, a predictive test used prior to selection of Air Traffic Control candidates, and the scores in the Initial En Route Qualification training course, a course designed to teach and measure actual Air Traffic Control job functions.

Research Hypotheses

H_1 There is a positive relationship between the composite score on the AT-SAT test battery and the composite score in the Initial En Route Qualification training course.

H_2 There is a positive relationship between the individual subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

H_3 There is a negative relationship between the individual subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

H_0 There is no clear relationship between the composite score or subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

Assumptions

For the purpose of this study, the following assumptions have been accepted:

1. Not all subjects are equal in intelligence quotient (IQ).
2. Not all subjects possess equal education levels.
3. Not all subjects respond to learning in the same manner or pace.
4. Some subjects may have learning disorders that were not reported or known.

Limitations

For the purpose of this study, the following limitations have been accepted:

1. Only students that attended training from January 18, 2011 through August 24, 2011 were chosen for this study.
2. Data collected on all students will be sanitized of any information concerning race, national origin, sex/gender/sexual orientation, age, or any other protected class information.
3. Only students that earned a score of 70% or greater on the Air Traffic Selection and Training (AT-SAT) test battery are selected to attend the Initial

En Route Qualification training course. Restriction of range results from this limited data source.

Definitions

The following definitions are furnished to provide common understanding of the terms used in this study:

Adverse Impact	The operational use of a selection instrument that results in negative consequences.
Air Route Traffic Control Center (ARTCC)	A facility that staffs air traffic controllers to manage the flow of traffic over large distances between an aircraft's departure airport and destination airport.
Air Traffic Selection and Training (AT-SAT)	A computerized test battery that measures skills, abilities, personal characteristics, and air traffic control knowledge that were validated through a comprehensive job analysis.
Controller Workforce Plan (CWP)	The FAA's strategy for hiring and training new air traffic controllers.
Cumulative testing	A testing strategy that builds on and retests all previous course material.
Developmental	A trainee that has not completed all required training.
eLearning	Electronically delivered teaching and learning products (Computer Based Instruction).

En Route Automation Modernization (ERAM)	The next generation of computer software that will provide all En Route data processing.
Front Line Manager (FLM)	An air traffic controller’s first level supervisor.
Instructional Systems Design (ISD)	A systematic method to develop training.
National Airspace System (NAS)	A term used to describe the entire aviation system in the United States.
Operational Tryout	A phase in the Instructional Systems Design in which the course is delivered to subjects for the first time.
Performance Verification (PV)	A pass/fail performance evaluation administered to students attending the FAA Academy from 1992-2011.
Professional Air Traffic Controllers Organization (PATCO)	The union that represented the air traffic controller workforce during the controller strike in 1981. President Ronald Reagan terminated the employment of 11,345 controllers.

Scope of the Study

This study focuses on the development and use of the AT-SAT test battery and the Initial En Route Qualification training course for the selection, training, and evaluation of air traffic controller candidates.

Organization of the Study

Chapter II is a review of the literature as it relates to AT-SAT and factors relevant to En Route air traffic control training. These include: staffing shortages, strike recovery, air traffic control screen (1981-1992), en route air traffic control training program (1992-2010), Air Traffic Selection and Training (AT-SAT) test battery, and en route air traffic control training program (2011 to present). Chapter III presents the study design, sampling technique, selection, development, and validation of the instruments, adverse impact, data gathering, and statistical methods used in assessing the results. Chapter IV documents the analysis of the data and the findings for each hypothesis. Chapter V is a summary of the study and recommendations offered for future research.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

This literature review summarizes the history and research associated with the en route air traffic controller selection and training process and the efforts the FAA has used to improve its process.

Every organization struggles with the costs associated with the selection of employees. New employees, even if they are already trained to do the job they were selected for, take time to offset the costs associated with recruitment, selection, orientation, and training. In the air traffic control profession, a new employee will not be fully certified for two to three years, making the selection of new employees very important. Training failures and turnover is expensive and is most common in newly hired employees (Wanous, 1992, p. 5).

Predicting the success prior to the hiring an individual can minimize the costs associated with the selection and training of new employees. Individuals that demonstrate a high aptitude begin at a higher level of performance and continue to learn at a faster rate than individuals that demonstrate a low aptitude for the job. The difference in aptitude level may not be great at the beginning of the training period, but these differences become more pronounced as the training proceeds (Drake, 1942, p. 102). Schmitt, et al., described predictor construct as an aspect of an individual that, if assessed, may predict future performance. Some examples are abilities, skills, or motivations (Schmitt, et al., 1993, p. 101).

A systematic process must be used when developing a program that predicts future success in an occupation. The job description is a logical place to start since this is the expected behaviors and actions that an employee should perform. A good job description will describe the results that should be achieved (Smith & Robertson, 1993, pp. 36-37). Another important factor is to identify the behaviors to be sought, define the conditions under which the behavior will occur, and define the level of acceptable performance (Popham, 1992, p. 67).

Most successful programs conduct a needs assessment. To be effective, the program must meet the needs of the participant. Ways to determine the needs are: Ask the participant, ask the participant's supervisor, ask others that are familiar with the job, test the participants, and analyze the performance appraisals of the participants. Interviews are very informative and may be supported or replaced by surveys (Kirkpatrick, 1998, p. 4). Simply stated, the desired status of learners minus the currents status of learners equals the educational need (Popham, 1992, p. 67).

The instrument used to test candidates can take several forms. Interviews are often used; however, in a technical profession such as air traffic control, the use of simulation can be very useful. Simulation in selection instruments are thought of as mini-replicas of the important tasks in a job. These simulations are an approximation of the job tasks, not a replica of the tasks. The most effective way to evaluate the effectiveness of the selection method is to correlate the scores from the method (simulation) to the subsequent measure of job performance (Wanus, 1992, pp. 126-127).

After the selection of the new employees, the orientation and training may begin. To evaluate the effectiveness of the training, five main concepts should be explored: 1) use a control group, 2) evaluate knowledge/skill before and after the training program, 3) use academic test to measure knowledge, 4) use performance test to measure skill, and 5) use the results of the evaluation to take the appropriate action (Kirkpatrick, 1998, p. 40).

The remainder of this chapter will be divided into the following sections: FAA air traffic controller staffing shortage, strike recovery, air traffic control screen (1981-1992), Air Traffic Selection and Training (AT-SAT) test battery, en route air traffic control training program (1992-2010), and en route air traffic control training program (2011 to present).

Section One: FAA Air Traffic Controller Staffing Shortage

The problem of air traffic controller staffing shortages has been well documented. In December 2004, the FAA published its first Controller Workforce Plan (CWP) that detailed a strategy to address staffing shortages in air traffic control facilities. An updated CWP was released in 2008. Both plans documented the plan to hire 17,000 new air traffic control specialists through 2017 (Krokos, et al., 2007, p. 1). According to the Department of Transportation – Office of Inspector General (2008), "A significant challenge for FAA will be training and certifying the large numbers of newly hired or "developmental[1]" controllers at their respective facilities; controllers can take up to 3 years to complete training" (p. 1).

[1] Developmental: An employee designated as a trainee, working towards full certification.

Dunleavy, et al. (2006), described the FAA's hiring policy as "one retirement, one hire". The training length for an en route air traffic controller is typically two to three years in length, with most of the training occurring on-the-job. A journeyman[2] air traffic controller is utilized for this training, taking valuable resources away from the daily operations. To maintaining the safety of the operations, the number of trainees must be limited to an acceptable number (p. 1).

The FAA has seen a much higher controller retirement rate than it had predicted. Since 2005, 3,300 controllers have left the workforce. This rate of attrition, which was 23 percent higher than the FAA had expected, resulted in an increase in controller hiring. From 2005 to 2008, 3,450 new air traffic controller candidates were hired, which was 25 percent more than had been projected. The consequence of these events is a dramatic increase in developmental controllers. In 2004, the developmental controllers accounted for 15 percent of the total workforce. By 2007, the number had increased to 25 percent (Department of Transportation – Office of Inspector General, 2008, p.2).

With the increased number of developmentals, the selection and training of these new controllers has become a high profile topic. A report identifying deficiencies in the screening[3], placement, and training of new air traffic controllers by the Department of Transportation Office of Inspector General (2010) added:

> Air traffic controllers play a critical role in maintaining the safety and efficiency of the National Airspace System. FAA continues to face a tremendous challenge in carrying out its goals to hire and train 15,000 new controllers over the next decade to replace those who were hired after the 1981 strike and are now retiring (p. 10).

[2] Journeyman: A fully certified controller that is now able to work without the assistance of an instructor.
[3] Screening: The process of selecting qualified candidates into the air traffic control field.

By 2017, 63 percent of the current air traffic controller workforce will become eligible to retire (GAO, 2008, p. 2).

Section Two: Strike Recovery

On August 3, 1981, the majority of the air traffic control workforce initiated a strike because of long-term labor unrest between the FAA and the Professional Air Traffic Controllers Organization (PATCO) (Department of Transportation Office of Inspector General, 2008, p. 1). President Ronald Reagan ordered the controllers back to work within 48 hours. Out of approximately 15,000 FAA air traffic controllers, 10,438 did not return to work. The employees that participated in the nationwide strike were fired and were banned from re-employment by a presidential directive (King, Manning, & Drechsler, 2007, pp. 1-2).

As a result of the strike and subsequent termination of the employees, the FAA began the mass hiring of replacement air traffic controllers. The time period between 1981 and 1992 became known as the strike recovery years in the FAA (GAO, 1997, p. 12). According to the Department of Transportation Office of Inspector General, 2008):

> To make up for the loss, FAA hired over 8,700 new controllers between 1982 and 1983, and, between 1983 and 1991, FAA hired an average of 2,655 controllers each year. By the end of fiscal year (FY) 1992, the strike recovery period had ended and controller hiring stabilized to the level of “one retirement—one hire.” This hiring wave created a large pool of controllers who have reached or will reach retirement eligibility at roughly the same time (p. 2).

The majority of the controllers hired during the strike recovery years have remained with the FAA and because they were required to be younger than 31 years of age at entry into the program, a large segment of the current workforce is or will be eligible for retirement (GAO, 1997, p. 12).

According to King, Manning, and Drechsler (2007):

> The post-strike hiring wave created the potential for a large portion of the controller workforce to reach retirement age at roughly the same time, particularly due to the FAA policy requiring retirement from controlling air traffic by age 56. Based on current projections, 73% of the agency's 15,000 controllers will become eligible to retire within ten years. Total losses are expected to reach nearly 11,000 (p. 1).

The Department of Transportation Office of Inspector General (2009) has reported that the FAA intends to select and train approximately 17,000 air traffic controllers in the next seven years to replace the numerous controllers hired during the strike recovery years that are now retiring (pp. 1-2). As King, Manning, and Drechsler have described, the large number of applicants required for this task has challenged the FAA to develop "effective recruitment, selection, and training procedures to ensure that it's staffing needs are met" (p.1).

Section Three: Air Traffic Control Screen (1981-1992)

In October of 1981, a new selection battery was implemented by the FAA. The two-stage selection process consisted of a paper and pencil test and a nine week FAA Academy screen program (Broach & Manning, 1997, p. 2). The initial test was administered by the Office of Personnel Management (OPM) and contained an abstract reasoning test. According to Broach and Manning (1997), the applicant was asked 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" (p. 2).

Applicants could also receive extra credit points by answering questions that demonstrated their knowledge of job-related subjects.

A list of candidates was established by ranking the scores of the OPM test combined with additional points given if the applicant is a veteran. The candidates with the highest scores were offered positions on the condition they could pass a medical and security clearance (Ramos, Heil, & Manning, 2001, pp. 1-2). The minimum qualification score was 70 percent; however, a score of 90 percent was normally required to be competitive with other applicants (Broach & Manning, 1997, p. 2). One drawback to the OPM testing was it had been used since 1981 without any major revisions. The test was compromised because of testing strategies and coaching programs offered through books and private companies. Over time, OPM test scores increased without a comparable increase in scores in the FAA Academy screen[4]. The artificial increase in OPM test scores reduced the effectiveness of the test to identify the candidates with the highest aptitude for the job. This resulted in a successful completion rate of less than 40 percent for air traffic control trainees during the nine week FAA Academy screen (Ramos, Heil, & Manning, 2001).

Candidates that were selected to attend the FAA Academy entered into a nine week screening process. The program was designed to assess the potential of the candidate by teaching nonradar air traffic control rules[5] and procedures and testing their ability to run laboratory scenarios (Broach & Manning, 1997. pp. 2-3). The students attended classroom sessions that taught aircraft characteristics, principles of flight, the

[4] FAA Academy screen: A nine week course that air traffic control candidates attended to determine their potential to become air traffic control specialist.
[5] Nonradar air traffic control rules: Air traffic control rules designed to safely move aircraft from one geographic area to another without the aid of radar surveillance.

national airspace system, and basic nonradar separation rules. This was followed by laboratory sessions that allowed the students to simulate the use of these rules and prepare for the laboratory evaluations (Ramos, Heil, & Manning, 2001, pp. 1-2).

Former air traffic controllers were hired and trained to evaluate the student's performance during the screen. During the nine week course, 13 items were given numerical scores. Classroom academics were weighted at 20 percent of the total grade, the laboratory scenarios were worth 60 percent, and a final comprehensive exam was weighted 20 percent. The laboratory scenarios were graded by using a two-part grading approach. First, the evaluator assessed the student on their technical ability by using a checklist to document the errors observed. The second part of the score was the evaluator's subjective opinion of the student's ability. Each part was worth 50 percent of the overall laboratory score. Students that accumulated a score of 70 percent or greater were assigned an air traffic control facility and continued training. Students that did not score 70 percent were removed from the air traffic control occupation (Broach & Manning, 1997, p. 3).

The process for selecting and screening air traffic control candidates received criticism in the late 1980s for two major reasons. The program was expensive and the OPM entrance exam process was being compromised (Bleckley, 2010, p. 4). The FAA committed $20 to 25 million each year to receive approximately 1,400 new trainees (Broach & Manning, 1997, pp. 3-4). Department of Transportation Office of Inspector General (2010) estimated the cost per applicant at $20,000 (pp. 2-3). The FAA costs were not the only issue. The air traffic control trainees also incurred significant costs. Most applicants that attended the FAA Academy screen resigned and took a leave of absence from their current jobs. Some applicants also had the expenses associated with leaving their families for nine weeks while only having a 55 percent chance of remaining with the FAA at the end of the training program (Broach and Manning, 1997, pp. 3-4).

Section Four: Air Traffic Selection and Training (AT-SAT) Test Battery

The development of an assessment tool to select air traffic controller candidates is a critical element in reducing costs associated with training failures in the FAA (King, Manning, & Drechsler, 2007). AT-SAT was developed to reduced costs, maintain validity, and reduce adverse impact on minorities and women through a computer selection battery that has a direct correlation to the job and is legally defensible (Ramos, Heil, & Manning, 2001, p.2; Broach & Manning, 1997, p. 4). An early decision to deliver the test via computer added cost savings through the administration, scoring, and data retrieval/storage (Kveton, et al., 2007).

The AT-SAT test battery was designed to predict the success of potential air traffic control applicants and to infer the aptitude of those candidates (Dunleavy, et al., 2006, p. 14; King, Manning, & Drechsler, 2006, p. 1). This development was based on the Separation and Control Hiring Assessment (SACHA) task analysis conducted in 1995 by Nickles, Bobko, Blair, Sands, and Tartak (King, Manning, & Drechsler, 2007, p. 1). Dunleavy, et al. (2006) explained how the initial design of the AT-SAT test battery was developed:

> ATCS characteristics were categorized into the following general categories: reasoning, computational ability, communication, attention, memory, meta-cognitive, information processing, perceptual abilities, spatial abilities, interpersonal, work and effort, stability/adjustment, self-efficacy, and psychomotor ability. These characteristics drove the development of various subtests of the AT-SAT, and as part of the original validation effort, subject matter experts made linkage judgments between AT-SAT subtests and the worker characteristics identified by the SACHA analysis.

It is important to remember that AT-SAT is an aptitude test, not a test designed to measure air traffic control knowledge (King, Manning, & Drechsler, 2007, pp. 1-2).

There are eight sub-tests that comprise the AT-SAT test battery. When administering the test, each sub-test begins with a comprehensive explanation of the test and ungraded questions or practice exercises (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 3). The following sub-sections will describe each sub-test in greater detail: Scan, dials, angles, applied math, analogies, letter factory, air traffic scenarios, experience questionnaire, and testing and validation of AT-SAT.

<u>Scan</u>

The scan test is designed to present the applicant with numerous data fields, called data blocks. These data blocks move at random speeds and directions across the computer screen. The data blocks contain two lines of data: an identifier on the top line and a three digit number on the bottom line. At the bottom of the computer monitor a range of numbers is displayed. Throughout the exercise, the range changes, prompting the test taker to respond to pick the data blocks that display a number within the currently displayed range (Tsacoumis, Anderson, & King, 2006, pp. VI-1-VI-2). An example of a scan test screen is provided in Figure 2-1 below (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 4).

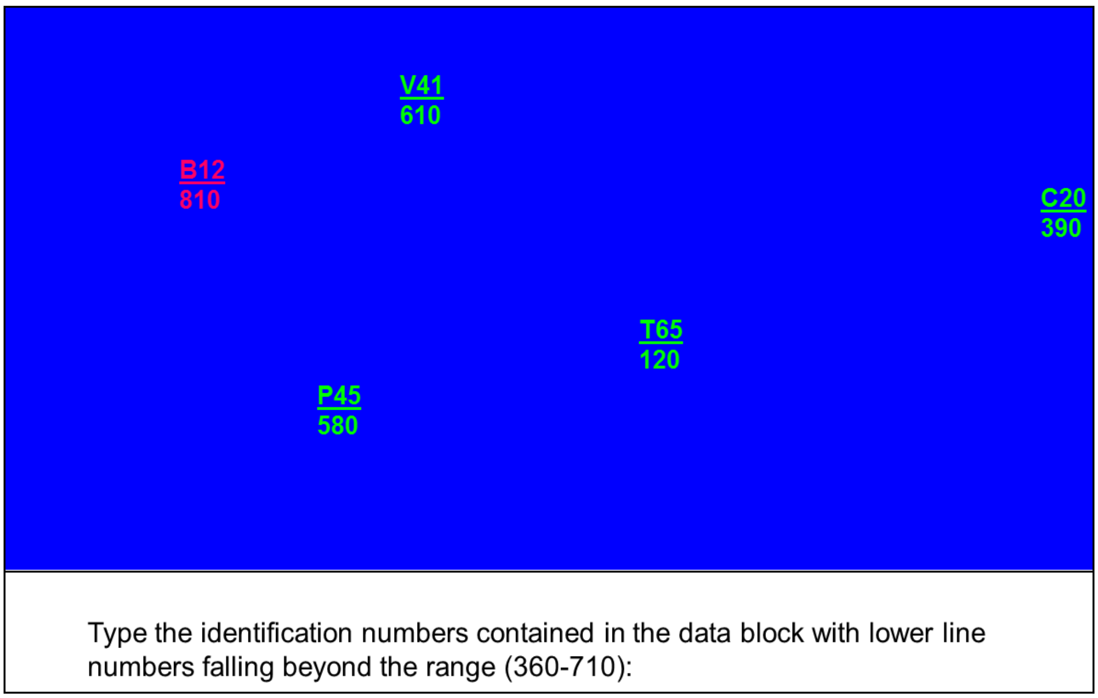

Figure 2-1. Example of the AT-SAT Scan Test Screen.

According to Tsacoumis, Anderson, and King (2006) the worker requirements assessed by this test are:

- Scanning – the ability to quickly and accurately search for information on a computer screen, radar scope, or computer printout
- Perceptual Speed and Accuracy – the ability to perceive visual information quickly and accurately and to perform simple processing tasks with it (e.g., comparisons)
- Dynamic Visual-Spatial – the ability to deal with dynamic visual movement (pp. VI-1-VI-2).

Dials

The dials test is designed to test the applicant's ability to interpret visual information and process that information into simple comparisons, or simply stated, the ability to read dials quickly and accurately (FAA Civil Aeromedical Institute Human Resources Research Division, n.d., p. 5). The applicant has nine minutes to respond to multiple screens that display an aircraft instrument panel. The panel contains two rows of seven gauges: Voltmeter, RPM, fuel-air-ratio, altitude, amperes, temperature, and airspeed (Dunleavy, et al., p. 14). An example of the dials test screen is provided below in Figure 2-2 (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 5).

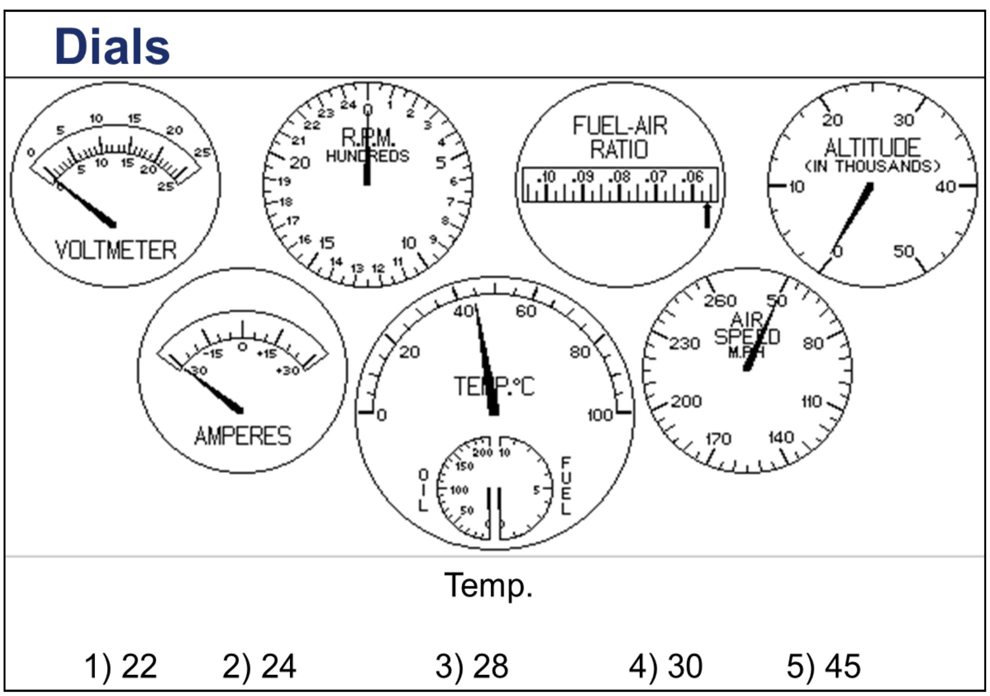

Figure 2-2. Example of the AT-SAT Dials Test Screen.

According to Tsacoumis, Anderson, and King (2006) the worker requirements assessed by this test are:

- Scanning – the ability to quickly and accurately search for information on a computer screen, radar scope, or computer printout
- Perceptual Speed and Accuracy – the ability to perceive visual information quickly and accurately and to perform simple processing tasks with it (e.g., comparisons) (p. IV-1).

Angles

The angles sub-test is designed to test the applicant's ability to interpret angles (FAA Civil Aeromedical Institute Human Resources Research Division, n.d., p. 6). The applicant has eight minutes to respond to 30 multiple choice questions. Each question has four response options (only one is correct). There are two types of questions. On the first 15 questions, the applicant is shown a graphic with an angle. The text asks the applicant to identify the measurement of the angle, for example, 20 degrees, 30 degrees, 40 degrees, or 50 degrees (Tsacoumis, Anderson, and King, 2006, p. II-1). An example of an angles test question, part 1, is provided below in Figure 2-3 (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 6).

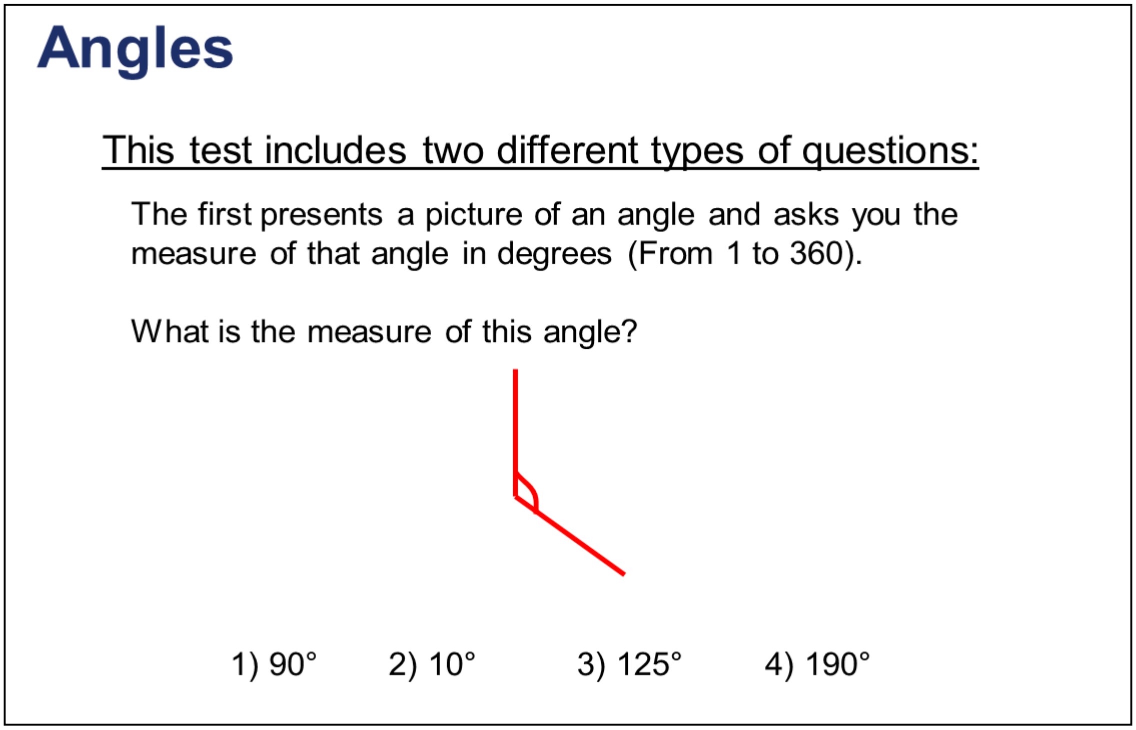

Figure 2-3. Example of AT-SAT Angles Test Question, Part 1.

In the second part of the angles test (the second 15 questions) the applicant is presented four angles and asked to identify which of four angles represents a specific degree (Tsacoumis, Anderson, and King, 2006, p. II-1).

According to Tsacoumis, Anderson, and King (2006) the worker requirements assessed by this test are:

- The ability to apply the principles of geometry to angles and computations involving angles
- Speed and accuracy of computation (p. II-1).

Applied Math

The applied math sub-test is designed to test the applicant's ability to factor time and distance through the use of word math problems. The questions contain information for speed, time, or distance calculations (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 7). The applicants have 21 minutes to complete 25 multiple choice questions. An example of an applied math test question is: An aircraft has flown for 2 hours with a ground speed of 240 knots. How far did the aircraft travel? (Dunleavy, et al., 2006). An example of the applied math test question is provided below in Figure 2-4 (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 7).

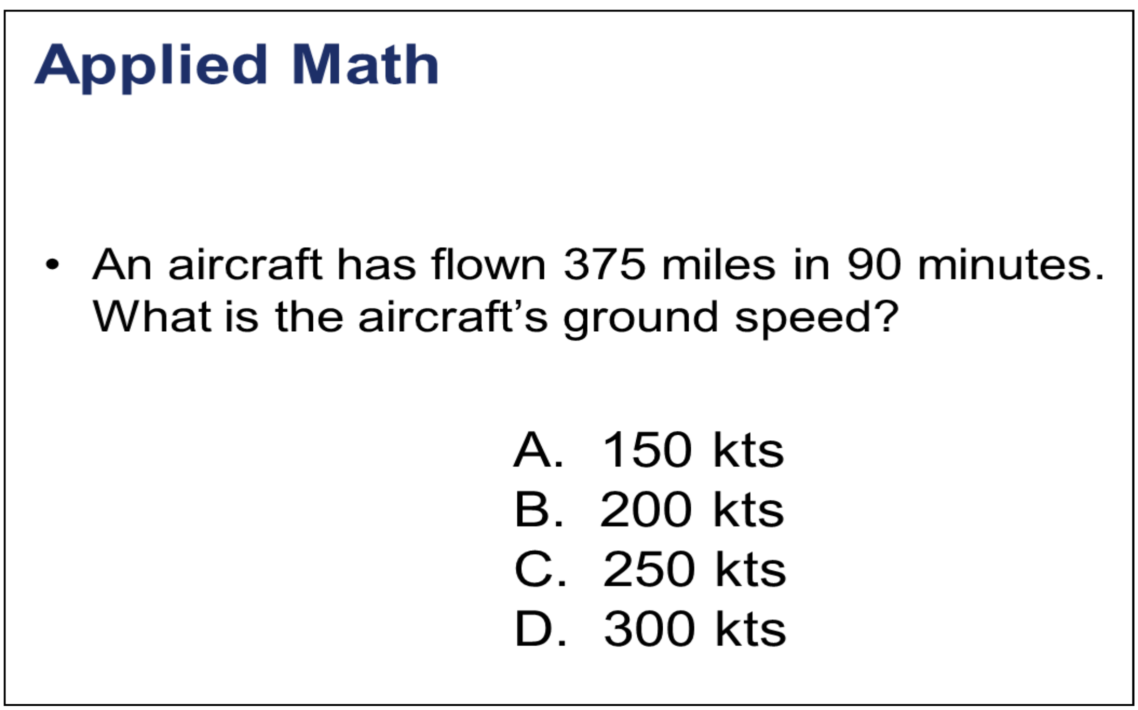

Figure 2-4. Example of AT-SAT Applied Math Test Question.

According to Tsacoumis, Anderson, and King (2006) the worker requirements assessed by this test are:

- Mathematical Reasoning
- Numeric Ability (addition/subtraction)
- Numeric Ability (multiplication/division) (p. VI-1).

Analogies

The analogies sub-test is designed to test the applicant's ability to solve reasoning problems through the use of visual and word problems. The applicants are asked to determine the relationship between words or figures in 46 multiple choice questions, 30 word problems and 16 visual items. An example of an analogy test word problem and visual item are provided below in Figure 2-5 and 2-6, respectively (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 9).

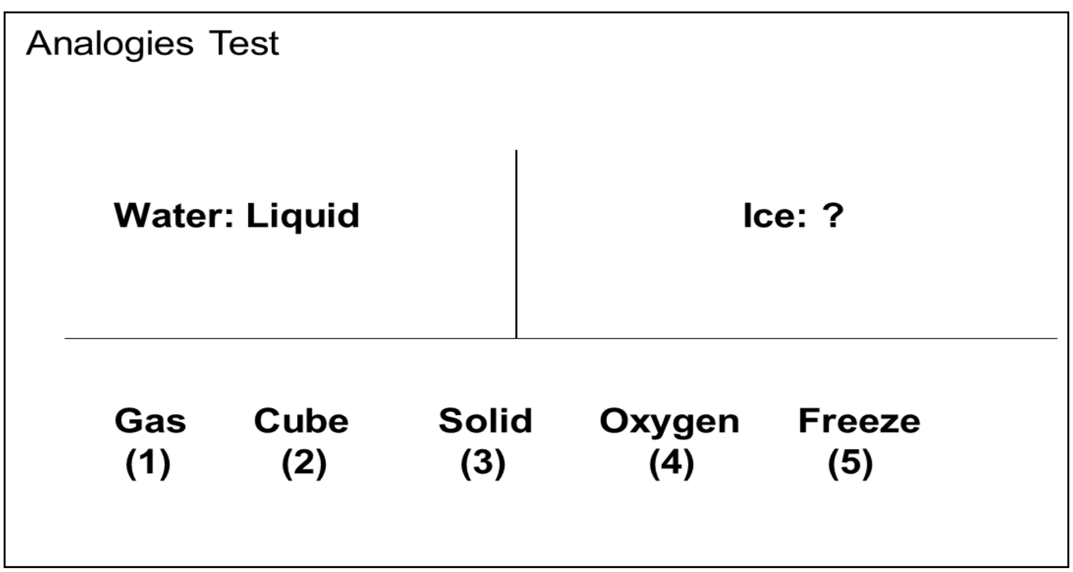

Figure 2-5. Example of AT-SAT Analogy (word problem) Test Question.

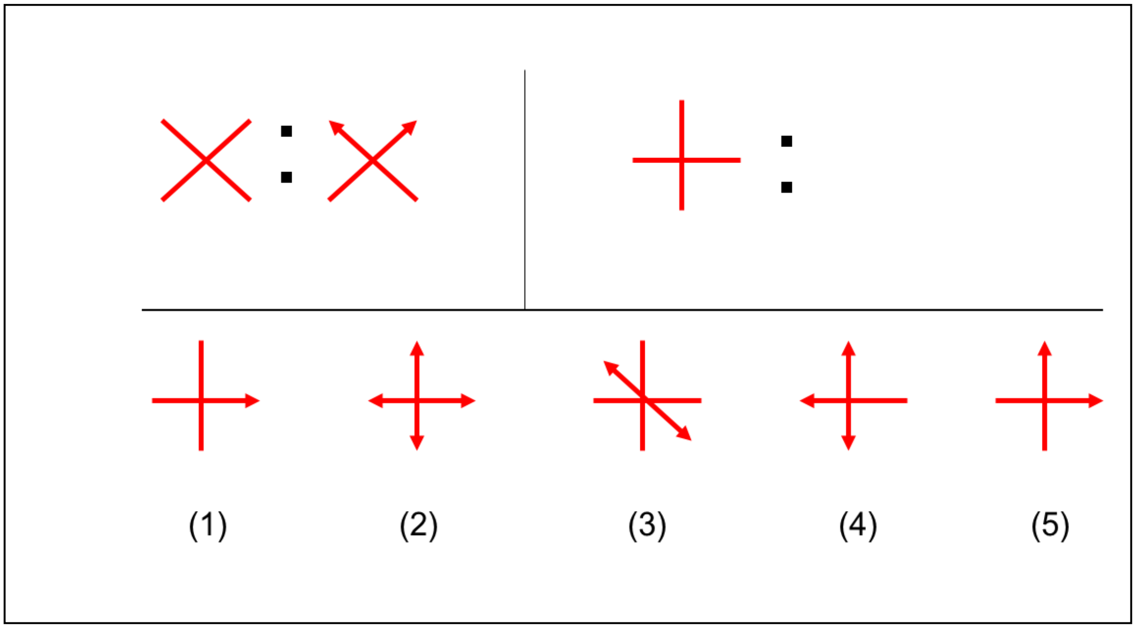

Figure 2-6. Example of AT-SAT Analogy (visual) Test Question.

According to Tsacoumis, Anderson, and King (2006) the worker requirements assessed by this test are:

- Reasoning - the ability to apply available information in order to make decisions, draw conclusions, or identify alternative solutions
- Visual-Spatial Reasoning - the ability to perceive and understand principles governing relationships among several figures
- Confirmation - the ability to efficiently select a response option consistent with the application of inferred rules
- Rule Inference - the ability to efficiently ascertain the rules governing relations between stimulus attributes
- Rule Application – the ability to efficiently apply transformational rules inferred from the complete portions of the stimulus array to the incomplete portion of the array (p. III-1).

Letter Factory

The letter factory test is designed to assess the applicant's ability to plan, think ahead, and maintain awareness. The test has two major parts: Letter factory conveyor and recall from memory.

The letter factory conveyor test was described by Porter, et al. (1996):

> The Letter Factory test presents four conveyor belts that move letters (A, B, C, and D) into one central area at varying speeds. Each letter may appear in one of three different colors: red, blue, or green. When a letter gets to the end of a belt, the examinee must place it into an empty box of the same color before it falls off of the belt. The examinee must fill each empty box with one A, one B, one C, and one D. The examinee must fill as many boxes as possible in the given time period. *Note:* examinees cannot control the distribution of letters or the speeds of the conveyor belts.
>
> Each scenario begins with an empty table and four empty conveyor belts. As the colored letters begin to appear on the conveyor belts, the examinee must determine the appropriate colored box to select from the group of empty boxes located on the left of the screen. For example, assume that of the first four letters to appear on the conveyor belt, letters A and B are blue and letters C and D are red. In this case, the examinee should place a blue box and a red box on the table. To place a box on the table, the examinee uses the mouse to "click" on the appropriate box and "drag" it to the table.
>
> As a letter approaches the end of a conveyor belt (after crossing the black line), the letter begins to blink. This indicates that the letter is "available" for placement into a box and that the examinee may "pick up" the letter (click the mouse on it and drag it to a box). The distance between the line and the end of a belt varies between belts. Also, note that several letters are available for placement at any one time (pp.2-3). To the right of the screen, there is an "energy gauge" that reports the energy level (in percentage) of the examinee at any given time. When the examinee's energy level falls below 25%, the "Energy Level" sign will turn red. The examinee must click on this sign as soon as it turns red. Then they can "take a break" any time before the energy gauge reaches the red line (by clicking on the appropriate icon). Instructions will inform examinees that a co-worker will relieve them for a break. During the examinee's break, the test screen will appear blank except for a message indicating when the test will resume or when the examinee will return from their break. Once they are refreshed (after a few seconds), the system will return the examinee to the workroom. Of course, the examinee must then become re-oriented with the new condition of the belts, letters, and boxes since things will have changed while they

> were out. If the examinee does not take a break before the gauge reaches the red line, they will be "too weak" to perform their job. If this occurs, letters will continue to become available and eventually fall off of the belt. It may be several seconds before the examinee regains control of the situation. If an examinee does not take a break and letters fall off the belt, the system will penalize the examinee for each letter that falls off the belt (pp. 5-6).

Several test sequences are presented to the applicant ranging from 30 seconds to two minutes 45 seconds. An example of a letter factory (conveyor) test screen is provided below in Figure 2-7 (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 10).

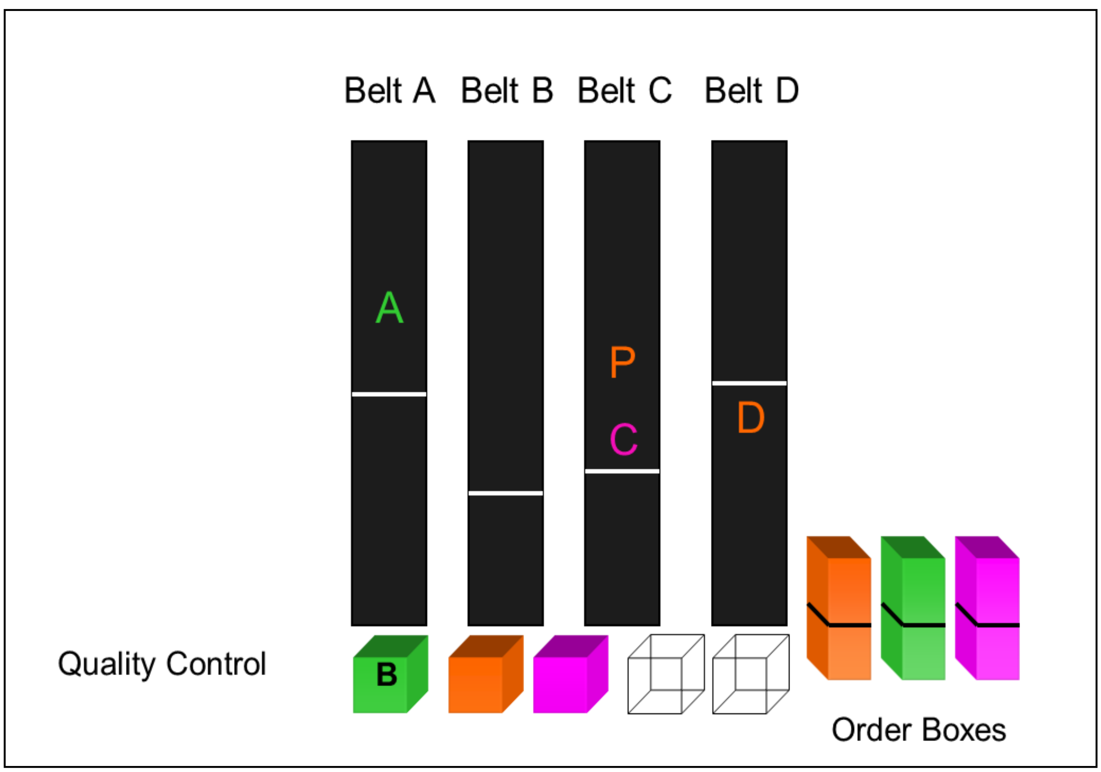

Figure 2-7. Example of AT-SAT Letter Factory (conveyor) Test Screen.

The letter factory recall from memory test is designed to measure the applicants recall from interruption and memory. Porter, et al. (1996) described the process as:

> Approximately three times during the test session, the test will pause and the screen will clear. The computer will then present a series of multiple choice questions regarding the examinee's present situation (e.g., “Approximately how many blue boxes do you have left?” or “How much energy do you have left?”). Other items will ask the examinee to rate the importance of several tasks (e.g., "Order more boxes," "Take a break," or "Fill boxes") at that moment in the test. Note that the order of importance of these tasks will change many times during the test (pp. 5-6).

An example question in the letter factory (memory) test is provided below in Figure 2-8 (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 10).

Letter Factory Test - Situation Awareness Questions

How many boxes should be in the loading area in order to correctly place all the letters?

1. One
2. Two
3. Three
4. Four

Figure 2-8. Example of AT-SAT Letter Factory (memory) Test Question.

According to Tsacoumis, Anderson, and King (2006) the worker requirements assessed by this test are:

- Timesharing/multitasking
- Tolerance for high intensity
- Situational awareness
- Planning
- Execution
- Prioritization
- Thinking ahead
- Decisiveness
- Immediate-term memory
- Short-term memory
- Scanning
- Concentration
- Perceptual speed and awareness
- Dynamic visual spatial
- Projection
- Attention to detail
- Recall from interruption
- Sustained attention (p. VII-1).

Air Traffic Scenarios

The air traffic scenarios test is designed test the applicant's ability to guide aircraft safely and efficiently. By using simple rules, the candidate's lack of air traffic control knowledge is not a factor (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 11). Tsacoumis, Anderson, and King (2006) developed a detailed description of the evaluation:

> The examinee's goal is to maintain separation and control of varying numbers of simulated aircraft (represented by symbols plus a data block) within their designated airspace as efficiently as possible. Simulated aircraft either pass through the airspace or land at one of two airports within the airspace. Each aircraft indicates its present heading, speed, and altitude via its data block. Separation and control are achieved by communicating with each of the aircraft. This is accomplished by using the computer mouse to click on the data block representing the aircraft and then clicking on instructions including heading,

> speed, or altitude. New aircraft in the subject's airspace have data blocks appear in white that turn green once the subject has communicated with (clicked on) them. Rules for handling aircraft are as follows: (1) maintain a designated separation distance between planes, (2) land designated aircraft at their proper airport and in the proper landing direction flying at the lowest altitude and lowest speed, (3) route aircraft passing through the airspace to their designated exit at the highest altitude and highest speed (p. VIII-1).

An example screen for the air traffic scenarios test can be seen below in Figure 2-9 (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 11).

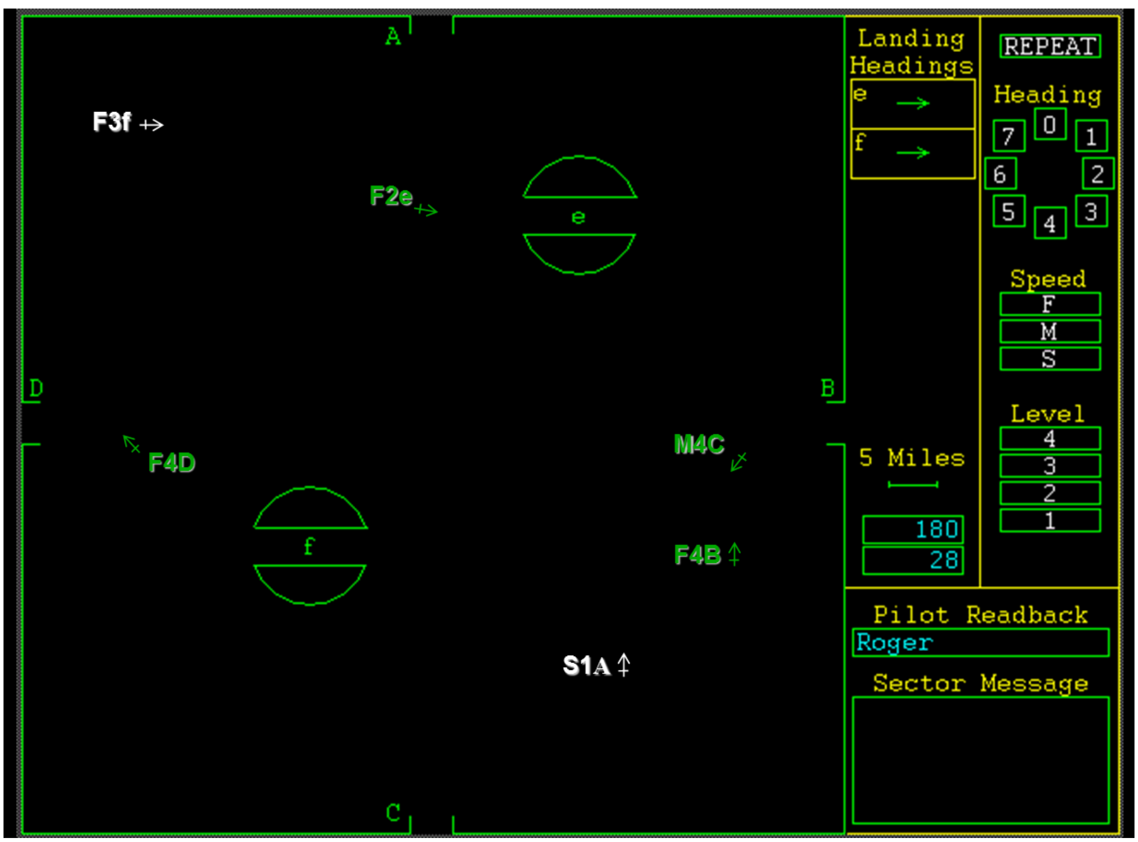

Figure 2-9. Example of AT-SAT Air Traffic Scenarios Test Screen.

According to Tsacoumis, Anderson, and King (2006) the worker requirements assessed by this test are:

- Prioritization
- Situational awareness
- Movement detection
- Tolerance for high intensity
- Planning
- Thinking ahead
- Execution
- Dynamic visual spatial
- Scanning
- Decisiveness
- Composer
- Reasoning
- Translating information (p. VIII-2).

Experience Questionnaire

The experience questionnaire is designed to quantify the applicants past experiences (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 8). These attributes are measured from the statements made from the applicant. Applicants are given 45 minutes to respond to 135 statements. As Tsacoumis, Anderson, and King (2006) described "Test-takers are asked to indicate the accuracy of the statement using a five-point scale (i.e., 1= definitely true, 2= somewhat true, 3= neither true nor false, 4= somewhat false, 5= definitely false)" (pp. IX-1-IX-2). An example of an experience question can be seen below in Figure 2-10 (FAA Civil Aeromedical Institute, Human Resources Research Division, n.d., p. 8).

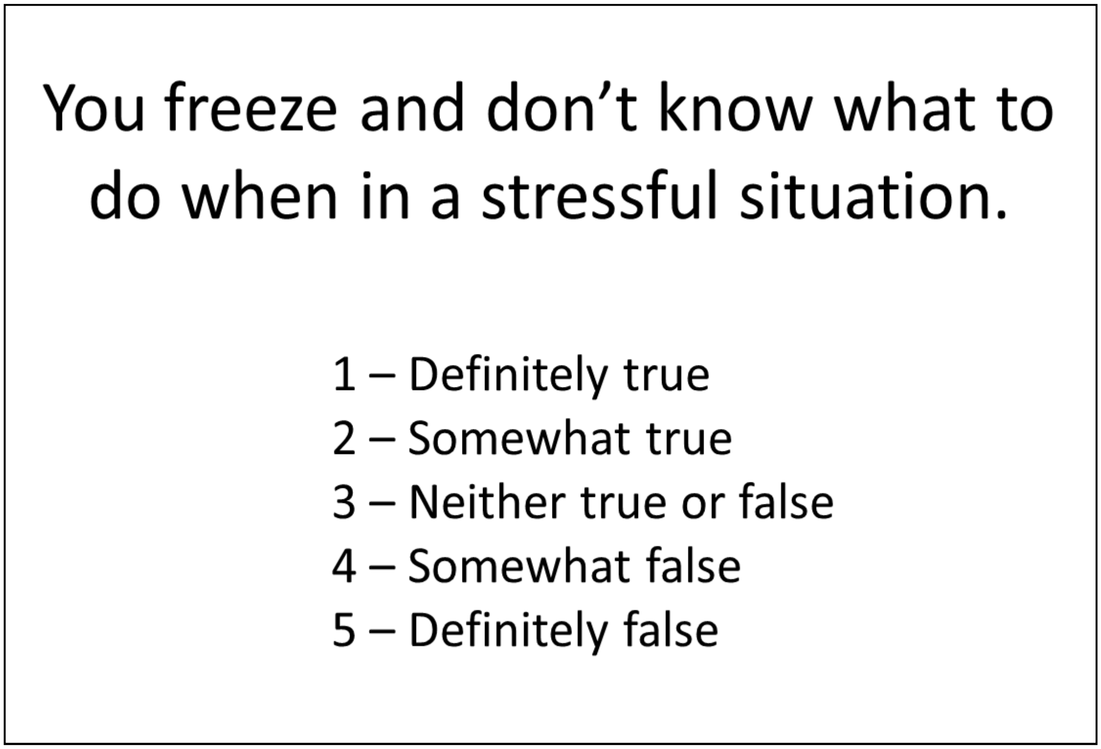

Figure 2-10. Example of AT-SAT Experience Question.

According to Tsacoumis, Anderson, and King (2006) the worker requirements assessed by this test are:

- Composure—the ability to think clearly in stressful situations
- Concentration—the ability to focus on job activities amid distractions for short periods of time
- Consistency of Work Behavior—the ability to behave consistently at work (e.g., dealing with coworkers in a consistent manner; consistently using the correct phraseology)
- Decisiveness—the ability to make effective decisions in a timely manner
- Execution—the ability to take timely action in order to avoid problems and to solve existing problems
- Flexibility—the ability to adapt to changing situations or conditions
- Interpersonal Tolerance—the ability to accommodate or deal with differences in personalities, criticisms, and interpersonal conflicts in the work environment

- Self-Awareness—the ability to maintain an internal awareness of one's actions and attitudes. This includes knowing one's limitations
- Self-Confidence—a belief that you are the person for the job and knowing that your processes and decisions are correct
- Sustained Attention—the ability to stay focused on a task(s) for long periods of time (over 60 minutes)
- Task Closure / Thoroughness—the ability to continue an activity to completion through the coordination and inspection of work (pp IX-1-IX-2).

Section Five: En Route Air Traffic Control Training Program (1992-2010)

The selection and initial training of air traffic control candidates was radically changed in the early 1990s. The following section is divided into three sub-sections: Air Traffic Selection and Training (AT-SAT), Initial En Route Qualification training course, and documented deficiencies of the En Route air traffic control training program (1992-2010).

Air Traffic Selection and Training (AT-SAT)

In 1990, the FAA began to revise the selection and training process for new air traffic control applicants. In a report by Broach and Manning (1997), the goals of the new selection process were to reduce the costs, maintain the validity, and reduce the adverse impact on women and minorities (p. 4). In 2002, a new screening tool, the Air Traffic Selection and Training (AT-SAT) test battery, was used for the first time. From 2002 to 2010, approximately 20,000 applicants have been tested as part of the selection process (Bleckley, 2010, p. 2).

The AT-SAT is a computerized test battery that measures skills, abilities, personal characteristics, and air traffic control knowledge that were validated through a comprehensive job analysis. An in-depth review on the AT-SAT and its validation method will be presented later in this chapter. Applicants that score a 70 percent or above on the AT-SAT are considered eligible for a job offer; however, very few applicants that score lower than 85 percent are selected because of the number of applicants with high scores. The AT-SAT is not used to determine the location or level of facility the applicants are assigned but many studies have been initiated to determine if this is a feasible option (Mitchell, 2010, p. 3).

Initial En Route Qualification Training Course

The most significant change to the FAA Academy training of new air traffic control applicants is that the program did not "screen" applicants. According to Ramos, Heil, and Manning (2001), because the applicants had been selected through the AT-SAT process, "the program will assume that candidates have the basic skills needed to perform the work of the ATCS" (p. 1). They added, "The nine-week-long screening process was replaced with a program that focused on training, rather than screening candidates for ATCS aptitudes" (p. 2).

The revised FAA Academy program was a four month course designed to teach students basic air traffic control rules and procedures. Many aspects of the previous FAA Academy screen were incorporated into the revised course, but the scope of the training was expanded to teach some radar concepts[6]. The most significant change in the program was the evaluation of the students. Only two evaluations, known as Performance Verification (PV), were administered during the training: an academic examination administered at the half-way point and a performance evaluation administered at the end of the course (Welp, et al., 2007, p. 4).

Performance Verification was administered by the Air Traffic Controller Training and Development Group. This office was a Washington Headquarters group located in Oklahoma City but not associated with the FAA Academy (Broach, 2009). This division of responsibilities allowed the FAA Academy to train the students without having the responsibility of also evaluating student performance. The academic evaluation was administered by the Air Traffic Controller Training and Development Group after the first half of the course was complete. Welp, et al. (2007) explained the process in an analysis of the course:

> The exam is computer-based and consists of 50 multiple choice questions. There are six variations of the test. Students are allotted two hours to complete the exam; however, it rarely takes longer than 40 minutes. Students are given their score immediately after completion of the exam and receive feedback on the questions that were missed. A score of 70% is required to pass. Students that do not pass the knowledge test are allowed to retake the test one day later. If they are unsuccessful on their second attempt, their employment is terminated with the approval of their respective Service Areas (p. 4).

The performance evaluation was also administered by the Air Traffic Controller Training and Development Group. These evaluations were conducted in the full fidelity radar laboratory by a manager from an en route operations facility. The evaluation was approximately 45 minutes long and was designed to test the student's ability to

[6] Radar concepts: The application of Radar rules and procedures to safely move aircraft from one geographic area to another with the aid of Radar.

demonstrate basic air traffic control skills. If the student was unsuccessful in the performance evaluation, the student received three additional instructional scenarios with an FAA Academy instructor to improve his/her performance. The student was allowed a retake of the evaluation the next working day. The retake was conducted by two evaluators, both from the Air Traffic Controller Training and Development Group. As Welp, et al. (2007) explained:

> The student is graded on the same criteria as the first assessment. At the conclusion of the scenario, the assessors compare notes and attempt to come to a consensus. If a consensus cannot be reached (one evaluator grades as a pass and the other grades as a fail), the student is graded as a pass. If both evaluators grade the student as a fail, the student's employment is terminated with the approval of their respective Service Areas (p. 4).

The success rate for Performance Verification was approximately 95% (Department of Transportation Office of Inspector General, 2010. pp. 8-9).

Documented Deficiencies of the En Route Air Traffic Control Training Program (1992-2010)

The decision to replace the OPM/FAA Academy screen has been scrutinized since its inception. Prior to the implementation of the plan to utilize AT-SAT, the Associate Director for Transportation Issues (U. S. General Accounting Office, 1991) testified, "Conceptually, FAA's new plan seems reasonable; however, it remains to be seen whether it will be a better way to screen, train, and place controller candidates" (pp. 10-11). The report documented the Associate Director's concerns that the new process replaced a nine week screening process with an eight hour computerized test battery, citing "FAA will have to identify candidate aptitudes in a much shorter time" (pp. 10-11).

The report also documented the Associate Director's concern that valuable training time could be wasted on applicants that did not have the aptitude to become air traffic controllers.

Several deficiencies in the FAA Academy training were documented through the Inspector General and the FAA's Air Traffic Controller Training and Development Group. The majority of the concerns were in the evaluation of the students, not the content or method of instruction. The performance verification assessment was a pass/fail evaluation, and limited the amount of information provided about student's performance. Managers were unable to get information about the students that described their strengths and weaknesses and how they could improve the performance of the student (Department of Transportation Office of Inspector General, 2010, pp. 8-9). Welp, et al. (2007) added, "The current PV process results in a Pass/Fail assessment only (no numeric score) that limits the FAA's ability to analyze the relationship between training performance and later job performance" (p. 5).

The Inspector General (Department of Transportation Office of Inspector General, 2010) examined previous Performance Verification evaluations and noted:

> There is no objective standard by which the PV is graded. Scores are based strictly on the subjective assessment of the designated examiners, which we found can vary extensively. For example, we found instances where candidates passed the PV with two loss of separation errors during the PV while other candidates failed the PV with no loss of separation errors.

These deficiencies, among numerous others, were considered and acted on when the FAA decided to revise the initial en route training course in 2007. The revisions to the new course will be discussed in greater detail in the next section titled "air traffic control training program (2011 to present)".

Section Six: En Route Air Traffic Control Training Program (2011 to Present)

The significant increase in the hiring of new air traffic controllers has strained many of the en route operations facilities. Some facilities have 100 trainees in their facilities (most facilities have 250 to 400 controllers). Given that the training program for en route air traffic controllers is two to three years, facilities are struggling to find resources to support the training of these developmentals (Welp, et al., 2007, p. 1).

Mitchell (2010) reported, “facility managers that DOT spoke with stated that candidates arrive after passing Academy training unprepared to begin facility training, indicating the need to restructure FAA’s testing and training procedures” (p. 2). The Department of Transportation recommended a redesign of the FAA Academy training program (pp. 1-3).

The Air Traffic Controller Training and Development Group conducted a front-end analysis[7] in 2007. According to Welp, Broach, and Kelley (2010), “Based on a front-end analysis conducted in 2007, it was determined that the initial en route training conducted at the FAA Academy was not meeting the expectations of the facility trainers and management and that a redesign was required” p. 1). The following sub-sections will describe the process of this redesign effort. These sub-sections are: Front-end analysis and redesign of the En Route Initial Qualification course.

[7] Front-end analysis: A needs assessment that includes surveys, interviews, and extensive data gathering to determine the strengths and weakness of the current training program/course.

Front-End Analysis

A front-end analysis was conducted in 2007 to determine the status of the current training program and gather recommendations for improvement. The front-end analysis resulted in several recommendations for improving training. Two major findings were: 1) the need to redesign the performance verification process and 2) the need to teach nonradar training. The recommendations and key findings are:

1. Ensure students can see the 3-D picture and have the necessary cognitive abilities to do the job.
2. Fix the "learn and dump" syndrome by improving student retention of basic knowledge and skills.
3. Continue to teach nonradar, but shorten the time spent on it.
4. Re-design the performance verification (PV) process.
5. Teach User Request Evaluation Tool (URET)[8] to all students.
6. Teach at least some radar position skills.
7. Delays between the time Stage I[9] training ends and Stage III[10] training begins may be resulting in significant skill and knowledge decay (Welp, et al, 2007, pp. 2-6).

Prior to the front-end analysis, the teaching of nonradar rules and procedures was controversial. Many FAA managers believed that nonradar should be eliminated from the program because of the limited amount of nonradar airspace in the United States. Proponents of teaching nonradar believed that it facilitated the three dimensional thinking needed to learn air traffic control. The front-end analysis showed overwhelming support for the continuation of teaching nonradar in initial en route training. Out of 21 facilities surveyed, 17 replied that nonradar should be taught at the application level and should

[8] User Request Evaluation Tool (URET): A software program that replaced paper flight progress strips and aided air traffic controllers by providing notification of possible aircraft conflictions.
[9] Stage I: The designation given to the Initial En Route Qualification training course.
[10] Stage III: Training administered at the field facilities by field personnel. This stage includes classroom, simulation, OJT, and certification on the radar associate position.

have consequences for poor performance (Welp, et al., 2007, pp. 2-6). Sethumadhavan and Durso (2009) reported, the use of nonradar "may serve as an important selection tool in assessing the performance of student controllers in radar environments. Performance during nonradar trials predicted final radar performance (i.e., collisions and landed aircraft count) independent of the predictive power of cognitive variables and above and beyond earlier radar training" (p. 21).

Redesign of the Initial En Route Qualification Training Course

Over a two year period, all course materials, exercises, part-task training, computer-based exercises, and full-fidelity scenarios were redesigned. A course design based on learning progression where a carefully planned set of sub-skills and academic subjects are sequenced as building blocks to ultimately achieve an instructional outcome (Popham, 2008, p. 24). The course structure can be seen below in Table 2-1, Initial En Route Qualification training course design (FAA Air Traffic Controller Academy Oversight Group, 2010, pp. 2-4).

Table 2-1. Initial En Route Qualification Training Course Design

Lesson Title
Facility Training Overview
Course overview
Aero Center Airspace
Radio and Interphone Procedures
Flight Progress Strips
Recording Clearances and Control Information
Forwarding Flight Plan and Control Information
Letters of Agreement
General Control
Board Management
IFR Clearances and Route Assignments
Departure Procedures
Altimeter Settings and Altitude Assignments
Holding Procedures
Arrival and Approach Procedures
Vertical Separation
MAP TEST
Longitudinal Separation
Lateral Separation
Initial Separation of Departures
METAR – Computer Based Instruction
BLOCK I TEST

Table 2.1 (Continued)

Block II
Block II Introduction
Nonradar Lab PRACTICE Scenarios 1-26
CONTROLLER KNOWLEDGE TEST #1
IET-41 Aircraft Characteristics - eLearning
NONRADAR GRADED LAB SCENARIO #1
NONRADAR GRADED LAB SCENARIO #2

Lesson Title
Radar Data Display
Beacon Code Assignment
Radar Identification
Radar Handoff and Point Out
Radar Separation and Safety Alerts
Radar Vectoring
Emergencies
Weather Hazards
AIRCRAFT CHARACTERISTICS TEST
Military Operation
Safety Culture
BLOCK III TEST
Position Relief Briefing
Team Responsibilities
Simulated Voice Switch Communication System (SVSCS)
Ghost Pilot
ERAM Decision Support Tool (EDST) II - CBI
CONTROLLER KNOWLEDGE TEST #2
Computer Equipment Message Entry Messaging I
Computer Messaging II

Table 2.1 (Continued)

COMPUTER SKILLS CHECKLIST
Scanning Awareness
Lab Guidelines
Lab Introduction

Block IV
RADAR Associate Practice lab Problems 1-48
RADAR ASSOCIATE GRADED LAB SCENARIO #1
RADAR ASSOCIATE GRADED LAB SCENARIO #2
RADAR ASSOCIATE GRADED LAB SCENARIO #3

The most substantial change to the revised course was the evaluation process. As discussed before, the evaluation process (performance verification) for students attending the FAA Academy was the most criticized topic in the front-End analysis. According to Welp, et al. (2007):

> The most common criticisms were (1) that it is a one-time event that does not take student performance throughout the course into account (i.e., there should be checks throughout the course), and (2) that the standards of performance are not clear, objective, and consistently applied among evaluators (p. 7).

Three recommendations to improve the process were repeated by almost every group interviewed: 1) use a numeric scoring system, 2) raise the standards, and 3) use a common pool of evaluators to ensure inter-rater reliability.

Two employees of the National Airspace System Human Factors Safety Research Lab, Civil Aerospace Medical Institute were interviewed to determine the actions required to develop the evaluation process. Dr. Manning and Dr. Broach reported the following requirements:

1. A stable cadre or pool of evaluators is used.
2. The need for operational personnel to conduct PV is eliminated or significantly reduced.
3. Evaluators are trained to achieve to a specific level of inter-rater agreement. Inter-rater agreement is periodically assessed, and if below an acceptable level, refresher training is provided.
4. The performance of students in PV is captured electronically (scenario events, student inputs, and voice tapes) and on paper at a level of detail sufficient to support recreation or replay of the scenario in case of dispute. These records are retained by [office] for at least [time].
5. A formal, numeric scoring method is used by evaluators based on specific behavioral criteria that assess mastery of the critical skills taught in the course and could be used for placement decisions or to conduct statistical analyses of the relationship between PV performance and field training or later job performance.
6. At least three independent assessments of student mastery over the specific skill set taught in the initial en route training course is made utilizing different scenarios to prevent the PV recommendation from being unduly influenced by a single uncommon incident of poor performance.
7. Multiple en route PV scenarios are used to that students are not able to anticipate and practice the exact scenario on which they will be tested. The scenarios have been empirically equated on complexity (Welp, et al., 2007, pp. 47-48).

Following the first delivery of the course and use of the new evaluation process, many revisions were made. These changes are described in chapter three. The lesson materials and laboratory scenarios were judged to be academically sound and only minor editorial changes were made.

Summary of Chapter II

This chapter discussed numerous topics related to the field of air traffic control selection and training. The current staffing level for air traffic controllers has the FAA's top managers concerned. The PATCO strike in 1981 has created problems for the FAA over the last 30 years and its effects are still being felt today. The rate of attrition has been slightly higher than the rate of replacement and most estimates predict the situation will not be improving. The FAA Academy screen training program was considered successful; however, the cost, high failure rate, and time involved prompted a redesign of the course in 1992.

The Air Traffic Selection and Training (AT-SAT) test battery has been used to select new air traffic controllers since 2002. The Initial En Route Qualification training course used until 2011 was considered a successful program; however, the Performance Verification process was scrutinized by multiple organizations. This, among numerous other factors prompted a complete redesign of the training program.

CHAPTER III

METHODOLOGY

Introduction

The purpose of this study was to determine the relationship between the AT-SAT test battery, a predictive test used prior to selection of Air Traffic Control candidates, and the scores in the Initial En Route Qualification training course, a course designed to teach and measure actual Air Traffic Control job functions. The hypotheses were derived through the use of a quantitative research design strategy, which is available in the review of the literature. A quasi-experimental design was used because the selection of subjects was not random (Wiggins & Stevens, 1999, p. 61). Data were collected from all students attending the FAA Academy Initial En Route Training course between January 18 and August 24, 2011, 2011. The hypotheses were:

H_1 There is a positive relationship between the composite score on the AT-SAT test battery and the composite score in the Initial En Route Qualification training course.

H_2 There is a positive relationship between the individual subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

H_3 There is a negative relationship between the individual subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

H_0 There is no clear relationship between the composite score or subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

Type of Design

This study uses a quantitative design. The main textbooks used in the research of quantitative design were: 1) Wiersma, W., (1995). *Research methods in education: An introduction* (6^{th} ed.), 2) Gay, L. R., & Airasian, P. W. (2003). *Educational research: Competencies for analysis and applications* (7th ed.), 3) Grimm, L. G., (1993). *Statistical applications for the behavioral sciences,* and 4) Triola, M. F. (2007), *Elementary statistics using EXCEL* (3^{rd} ed.).

Selection of the Subjects

The entire population of students that attended the FAA's Initial En Route Qualification Training course in Oklahoma City, Oklahoma between January 18 and August 24, 2011 were included in this study. These students represented the first 156 students to complete the newly redesigned course. Data collected from each subject was confidentially stored within the memory of a laptop computer. In accordance with a research agreement with the Federal Aviation Administration's Executive Director for Human Resources Programs and Policies, the Civil Aerospace Medical Institute

(CAMI),and the Oklahoma State University Institutional Review Board, all data has been de-identified and is free of any information concerning race, national origin, sex/gender/sexual orientation, age, or any other protected class categories. A copy of these agreements can be found in Appendix A.

Selection/Development of the Instruments

Testing and Validation of AT-SAT

Accurately predicting future performance of a candidate is difficult, particularly when the applicant has no linking experience to the future job duties. This problem is not unique to the United States. Worldwide, optimizing the selection of air traffic controllers has been attempted (Oprins, Burggraaff, & Weerdenburg, 2006). A detailed analysis was performed by the FAA that identified the tasks performed by air traffic controllers and the specific skills to measure in support of those tasks (Department of Transportation Office of Inspector General, 2010, p. 1). The worker requirements (tasks, knowledges, skills, and abilities) of the air traffic control occupation were identified through a job analysis (Ramos, Heil, and Manning, 2001a). According to Nickels, et al. (1995):

> The foundation for building valid selection procedures consists of defining the job and identifying the skills and abilities that are required to perform the job. A method used to identify these job activities and worker requirements is job analysis, a systematic process for analyzing the tasks performed (both cognitive and behavioral) on a job and for determining the worker requirements (skills, abilities, and other personal attributes) which are necessary to perform those tasks (p. 21).

The Separation and Control Hiring Assessment (SACHA) project was developed as a foundation to build a selection procedure that was valid, legally defensible, and identified the skills and abilities to perform as an air traffic controller (Nickels, et al., 1995,

p. 21). The testing and validation process described in the following pages was directly related to the job analysis as described in the SACHA report.

The testing and validation of the AT-SAT test battery was accomplished over several years, with thousands of pages documenting the process. A summary of the development and validation can be found in Appendix C.

The final AT-SAT battery was selected based on several factors. One major factor was the amount of time required for the exam. The goals for elimination, according to Ramos, Heil, and Manning (2001b), were:

> 1. Maintain high concurrent validity.
> 2. Limit the test administration time to a reasonable amount.
> 3. Reduce differences between gender/racial group means.
> 4. No significant differences in prediction equations (i.e., regression slopes or intercepts) favoring males or whites (i.e., no unfairness).
> 5. Retain enough tests to allow the possibility of increasing the predictive validity as data becomes available in the future (pp. 37-41).

The final AT-SAT battery consisted of: scan, dials, angles, applied math, analogies, letter factory, air traffic scenarios, and experience questionnaire.

The validity of the final AT-SAT battery was estimated to be .76. This is an extremely high value; however, the High-Fidelity Performance Measure (HiFi) sample size was very low, impacting an accurate estimate for validity. As documented by Ramos, Heil, and Manning (2001b), "The most relevant validity of .76 is the correlation with the composite criterion which is corrected for range restriction, shrinkage, and criterion unreliability" (pp. 41-42). They authors added:

> The high correlations of the CBPM and Ratings with the high fidelity criteria are strong evidence that the CBPM and Ratings are accurate indicators of job performance. Inter-rater agreement reliability was used to correct the validities for the Ratings and HiFi criteria.

AT-SATs Relationship to Performance Verification

A longitudinal validity analysis conducted by Bleckley (2008) assessed the ability of the AT-SAT to correctly predict the success of a student in the Initial En Route Qualification training course (pre-2011 course evaluated by the performance verification process). Two characteristics of this evaluation were: 1) students that pass the AT-SAT (score 70 percent of above) are further categorized as “well qualified”, meaning they scored 85 percent or above, and 2) the Performance Verification scores are binary, having only a score of “pass” or “fail” (p. 1).

From a population of 650 students, 93 percent of the students that were “well qualified” on the AT-SAT battery passed their Performance Verification evaluation on their first attempt. Students that are not successful on their first attempt are given a second evaluation. Bleckley (2008) added, “There is a trend that suggests that the higher the AT-SAT score the more likely you are to pass PV on first attempt” (pp. 1-2).

Testing and Validation of the Initial En Route
Qualification Training Course

As previously described in Chapter II, The Initial En Route Qualification course was revised as a result of the Front-end analysis conducted in 2007. A team of headquarters specialists, instructional system designers, facility training managers, FAA Academy instructors, and contract training support specialist collaborated over a 3 year period to develop, evaluate, and validate the new course. The process was extensive and underwent numerous review/edit cycles. The process used to develop and validate the course can be found in Appendix D.

Adjustments of AT-SAT to Avoid Adverse Impact

The operational use of a selection procedure or instrument that results in negative consequences is referred to as adverse impact. Broach and Heil (1997) described:

> Adverse impact occurs when the rate, numbers, or proportions of protected persons selected for the job are statistically less than the rate, numbers, or proportions of the majority persons selected for the job on the basis of a selection procedure (p. 2).

The *Uniform Guidelines on Employee Selection Procedures,* 29 CFR 1607.4 (1978b), provides a rule of thumb that gave further explanation:

> A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact.

If adverse impact is suspected or proven, the *Uniform Guidelines on Employee Selection Procedures,* 29 CFR 1607.14 (1978a) suggests an investigation into the fairness of the procedure or instrument should be conducted by stating:

> Where a selection procedure results in an adverse impact on a race, sex, or ethnic group identified in accordance with the classifications set forth in section 4 above and that group is a significant factor in the relevant labor market, the user generally should investigate the possible existence of unfairness for that group if it is technically feasible to do so. The greater the severity of the adverse impact on a group, the greater the need to investigate the possible existence of unfairness.

The unfairness of the procedure or instrument is described by the *Uniform Guidelines on Employee Selection Procedures,* 29 CFR 1607.14 (1978a) as:

> When members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains lower scores.

In 1996, the AT-SAT test battery was evaluated for its potential as a selection test for future air traffic controllers. The test battery was weighed against the Uniform Guidelines for Employee Selection Procedures, as described above (Broach, 1996, p. 2). Prior to using the AT-SAT for the FAAs selection instrument, some minority groups were concerned about adverse impact. In particular, some special interest groups had raised concerns because only three percent of African American applicants were predicted to achieve the minimum passing score of 70 percent (King, Manning, & Drechsler, 2007, pp. 2-3). Dunleavy (2006) described in greater detail:

> For some of the subtests that were cognitive in nature, the 4/5ths rule was violated for blacks, Hispanics, and females. Mean differences were around -0.75 in some cases for cognitive subtests, and these results were a concern to FAA management. However, moderated multiple regression fairness analyses suggested that regression slopes across all the subgroups of interest were similar. Thus, the AT-SAT battery was not differentially valid across subgroups.

Despite evidence that the AT-SAT was not differentially valid across subgroups (and therefore did not demonstrate adverse impact), the FAA decided to re-weight the subtests.

The data from 724 developmental air traffic controllers were compared to determine the likelihood of mitigating the adverse impact by re-weighting the sub-test scores. The revised scores were calculated using the new weighting system and the mean score of all applicants increased 4.86 points; however, American Indian/Alaska Native, Hispanic, and African American applicants showed the greatest increase in mean scores (6.97, 6.98, and 7.02 points, respectively). The study showed that all groups resulted with higher scores and the potential for adverse impact was reduced (Dattel & King, 2006, pp. 1-6).

Data Gathering

AT-SAT scores were collected from the Civil Aerospace Medical Institute (CAMI). Each subject's composite score and individual sub-test competency score was entered into a spreadsheet and arranged by highest composite score to lowest composite score. The total composite AT-SAT score is derived from a formula that weights each of the 23 sub-test competencies. The formula and the raw data collected for each sub-test competency is not included in this study due to concerns of compromising the test by releasing this data to the general public.

The composite Initial En Route Qualification training course scores were collected from the FAA Academy. The data were consolidated with the matching AT-SAT data and stored on a personal laptop computer. All identifying data were removed to maintain the confidentiality of each student.

Reliability and Validation

Having laid the groundwork in the previous sections of this chapter, Content and Construct validity of the course were judged by members of the FAA Air Traffic Controller Training and Development Group, the Civil Aerospace Medical Institute, and the FAA Academy. The course content was determined to be consistent by all members. Criterion-related validity was evaluated throughout the design, development, review, and course delivery process. Each member of the review process was knowledgeable in the FAA lesson plan development and testing process.

Statistical Analysis

Two statistical analyses using correlation and regression study techniques were used in this study. The Pearson product moment correlation coefficient, also known as linear correlation coefficient r, was used to measure the linear relationship between the two quantitative variables, the AT-SAT composite scores and the composite scores of the Initial En Route Qualification training course (Triola, 2007, pp. 547-551). Multiple regression was used to determine the relationship between a response variable, the Initial En Route Qualification Training course, and multiple predictor variables, the AT-SAT test battery sub-tests (Triola, 2007, pp. 601-603).

Each research hypothesis presented unique properties that required the use of differing statistical analysis. In hypotheses one, linear correlation was used to determine the relationship between the AT-SAT test battery composite score (independent variable) and the Initial En Route Qualification training course composite score (dependent variable). Hypothesis two and three also used linear correlation to determine the relationship between each individual AT-SAT test battery composite score (independent variable) and the Initial En Route Qualification training course composite score (dependent variable).

Multiple regression was used in hypothesis two to determine which combination of AT-SAT test battery competencies presented the strongest relationship to the Initial En Route Qualification training course composite score. Regression was also used in the null hypothesis to determine if any excluded variables (AT-SAT test battery competencies) did not have a relationship to the Initial En Route Qualification training course composite score.

Summary

Chapter III described the process used to select the subjects and instruments for this study. The development and validation of the AT-SAT test battery and the Initial En Route Qualification training course was described in great detail. Both programs experienced numerous revisions prior to implementation. The method of gathering data and the statistical analysis used in the study were explained. Chapter IV will present the findings derived from the data collected from the AT-SAT test battery and the Initial En Route Qualification training course.

CHAPTER IV

FINDINGS

Introduction

This chapter presents the findings derived from the data collected from the AT-SAT test battery and the Initial En Route Qualification training course. The Initial En Route Qualification training course began conducting the redesigned curriculum on January 18, 2011. The course was deemed to be valid by the Air Traffic Controller Development and Training group and by research psychologists at the Civil Aerospace Medical Institute. The students attending this course were selected, in part, by their AT-SAT test scores. AT-SAT has been used as the primary selection tool for the FAA's air traffic controller workforce and has also been validated by the FAA. The development and validation of the course and the AT-SAT has been extensively described in the previous chapters.

After 156 students completed the Initial En Route Qualification training course, data about these students were collected. 18 students did not take the AT-SAT because they were hired as ex-military controllers and one student resigned prior to finishing the course. 137 subjects remained in this study (N=137). The raw data concerning all students is not included with this study. Any inquiries to receive/analyze this data must be requested through the FAA's Civil Aerospace Medical Institute, Oklahoma City, OK.

Demographic data for subjects are not provided. In accordance with the research agreement with the Federal Aviation Administration (see Appendix A), all demographics of the subjects in the study have been removed. Therefore, data regarding gender, education, work experience, etc. was not available for research. This chapter will review the raw data retrieved and document the relationship among the variables.

AT-SAT Data

The mean score for the AT-SAT battery was 91.31%, with a standard deviation of 5.65. Only students that passed the AT-SAT battery with a 70% or greater are selected for employment. This has resulted in a restriction of range. The lowest score in this study was 70.73% and the highest score 100%, a range of 29.27. Candidates for employment are categorized as “well qualified” if they score 85% or greater and have priority for employment. Of the 137 subjects in this study, 124 scored above 85%. The restriction of range issue is amplified since 90.51% of the subjects in this study scored 85% or greater on the AT-SAT battery.

As described in chapter two, the AT-SAT test battery consists of eight subtests. These subtests measure 23 separate competencies. These competencies are: Dials, applied math, scanning, angles, letter factory (planning), letter factory (awareness), Air Traffic Scenarios Test (efficiency), Air Traffic Scenarios Test (safety), Air Traffic Scenarios Test (accuracy), analogy, analogy (wind), analogy (latency), experience questionnaire (composure), experience questionnaire (consistency), experience questionnaire (concentration), experience questionnaire (decisiveness), experience questionnaire (self-confidence), experience questionnaire (tolerance), experience questionnaire (execution), experience questionnaire (thoroughness), experience questionnaire (flexibility), experience questionnaire (self-awareness), and experience questionnaire (sustained attention).

Initial En Route Qualification Training Course

The mean score for the Initial En Route Qualification training course was 79.58%, with a standard deviation of 10.11. The highest score was 98.36 and the lowest score was 44.06, resulting in a range of 54.30.

The Initial En Route Qualification training course pass rate for all subjects in this study was 85.40% (117 subjects out of 137 passed the course). As mentioned in the previous section, some students were categorized as “well qualified” by scoring 85% or greater on the AT-SAT test battery. These subjects’ mean Initial En Route Qualification training course score was 80.11% with a passing rate of 86.29% (107 subjects out of 124 passed the course). The mean Initial En Route Qualification training course score for subjects that were categorized as “qualified” (scored 70-84.99%) from the AT-SAT test battery was 74.50% with a passing rate was 76.92% (10 subjects out of 13 passed the course); however, caution should be taken prior to making any conclusions because of the limited amount of data available from the subjects that were “qualified”.

AT-SAT Test Battery Comparison to Initial En Route Qualification Training Course Scores

The mean composite scores from the AT-SAT test battery and the Initial En Route Qualification training course were 91.31% and 79.58%, respectively. The range, minimum score, maximum score, mean, and standard deviation of each AT-SAT battery competency is provided in Table 4-1, Descriptive statistics for the AT-SAT sub-test competencies, below.

Table 4-1. Descriptive Statistics for the AT-SAT Sub-Test Competencies

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Dials	137	7.000	13.000	20.000	19.00730	1.280489
Applied Math	137	16.000	9.000	25.000	20.71533	3.670128
Scanning	137	94.000	108.000	202.000	184.62774	11.674512
Angles	137	16.000	14.000	30.000	27.33577	2.544537
Letter factory (awareness)	137	38.000	26.000	64.000	45.13139	7.667557
Letter factory (planning)	137	.459	.078	.537	.27708	.072210
ATSC (Efficiency)	137	63.333	36.333	99.667	69.04866	15.271377
ATST (Safety)	137	81.333	18.667	100.000	66.95134	17.083157
ATST (Procedures)	137	73.333	20.333	93.667	54.63990	15.175664
Analogies	137	28.000	13.000	41.000	28.90511	5.945690
Analogy (Wind)	137	21.000	.000	21.000	6.72993	3.980658
Analogy (Latency)	137	11.821	12.821	24.642	17.77691	1.825758
Experience (Composure)	137	56.667	43.333	100.000	82.03163	10.678982
Experience (Consistency)	137	42.222	57.778	100.000	86.66667	8.678056
Experience (Concentration)	137	44.000	56.000	100.000	86.89051	10.085347
Experience (Decisiveness)	137	43.077	56.923	100.000	84.80629	9.770149
Experience (Confidence)	137	48.889	51.111	100.000	87.70479	10.036430
Experience (Tolerance)	137	34.000	66.000	100.000	89.98540	7.804585
Experience (Execution)	137	38.000	62.000	100.000	86.39416	9.480132
Experience (Thoroughness)	137	41.667	58.333	100.000	84.14842	9.879000
Experience (Flexibility)	137	49.333	50.667	100.000	86.62774	9.319235
Experience (Awareness)	137	44.000	56.000	100.000	83.98540	9.472860
Experience (Attentiveness)	137	50.000	50.000	100.000	86.56934	10.723665

The AT-SAT test battery and Initial En Route Qualification training course test scores can be found in Figure 4-1 below.

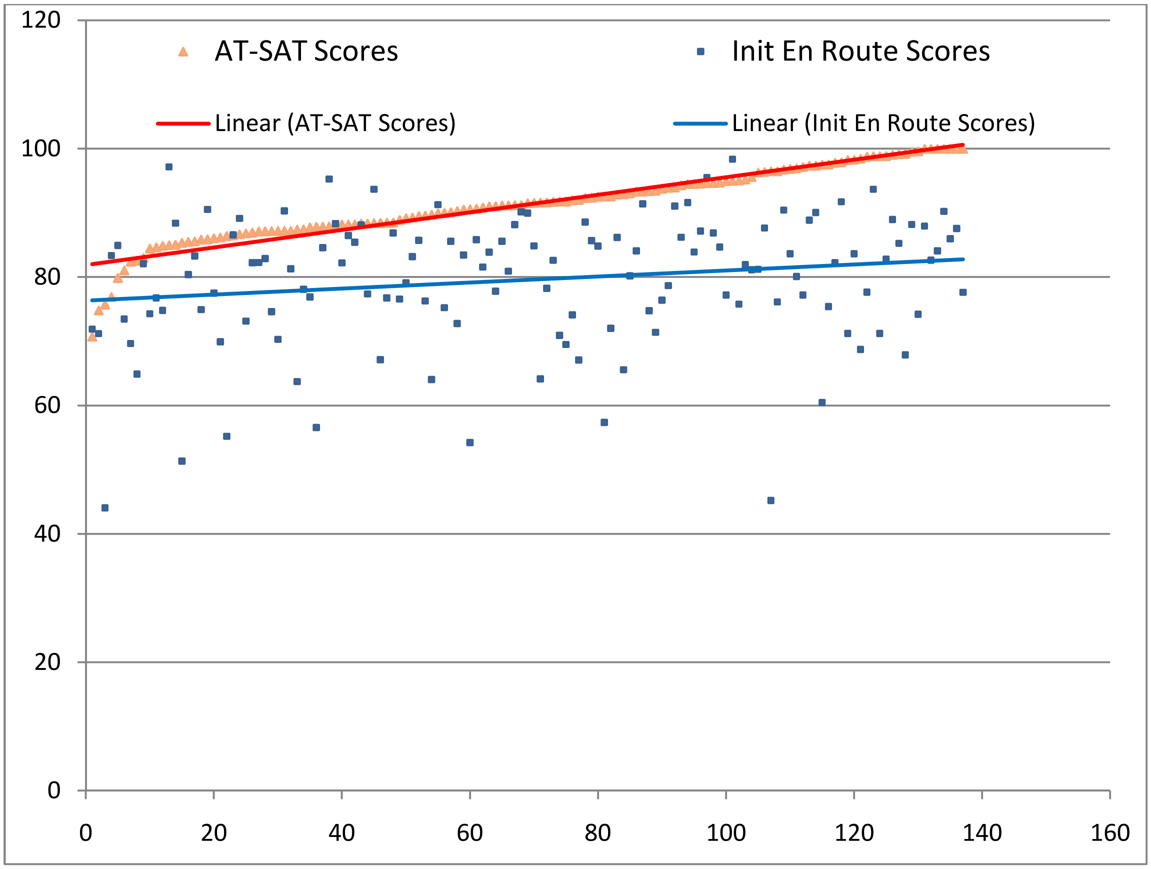

Figure 4-1. AT-SAT Battery and Initial En Route Qualification Course Test Scores.

<u>Hypothesis One</u>

H_1 There is a positive relationship between the composite score on the AT-SAT test battery and the composite score in the Initial En Route Qualification training course.

By using a bivariate correlation, a correlation coefficient of .216 was observed. The correlation was significant at the .05 level (two-tailed test of significance). The shared variance (R^2) between the AT-SAT composite score and the Initial En Route Qualification training course score was .047, or approximately 5%. A Venn diagram depicting the common variance between the AT-SAT composite score and the Initial En Route Qualification training course score can be found in Figure 4-2 below.

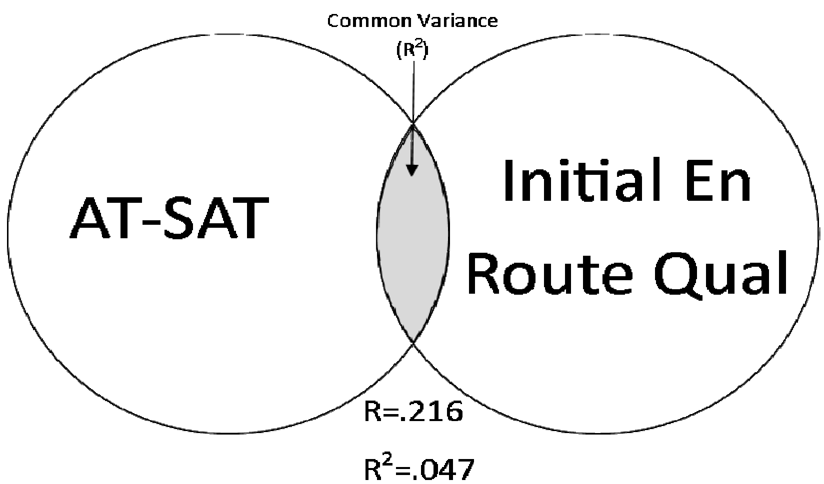

Figure 4-2. Shared variance (R^2) Between the AT-SAT Composite Score and the Initial En Route Qualification Training Course Score

The relationship between the composite score on the AT-SAT test battery and the composite score on the Initial En Route Qualification training course is a moderate positive relationship. Relation or correlation results rarely exceed .500 in the behavioral sciences (Cohen, 1977, p. 284; Quinnipiac, n.d.). The scatterplot can be seen in Figure 4-3 below.

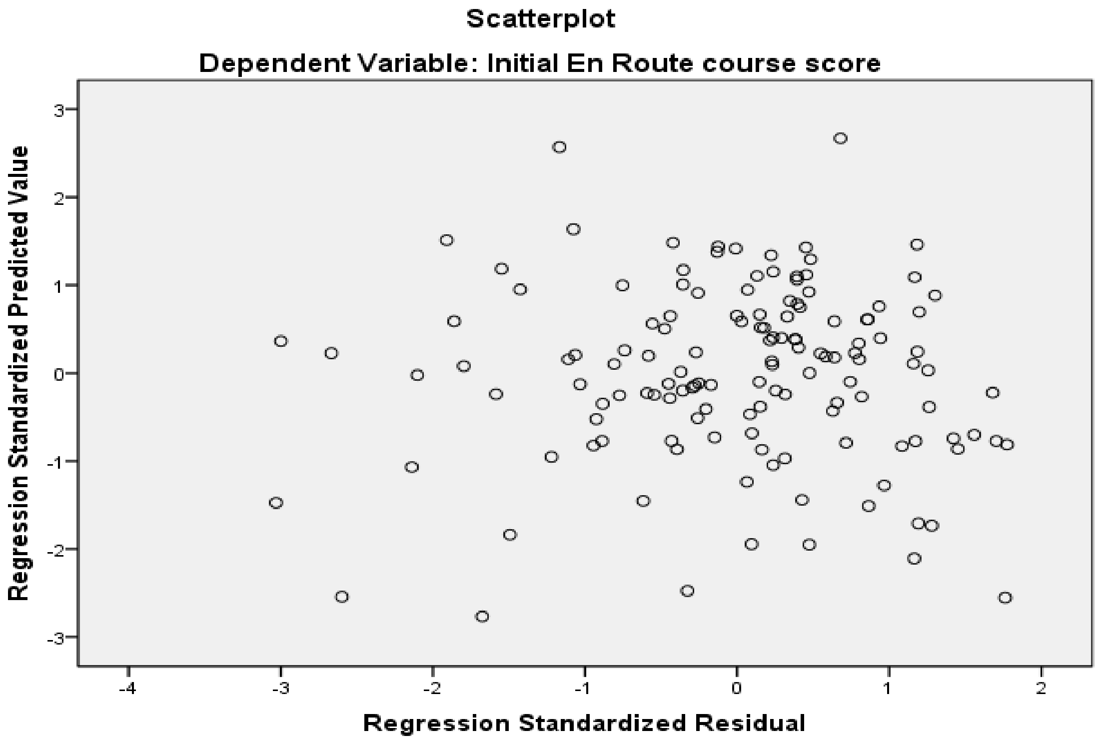

Figure 4-3. Scatterplot of the Correlation Between the Composite AT-SAT Test Battery Scores and Composite Initial En Route Qualification Training Course Scores.

<u>Hypothesis Two</u>

H_2 There is a positive relationship between the individual subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

A relationship of the individual competencies to the Initial En Route Qualification training course was determined by using bivariate correlation. The competencies that were determined to have a positive relationship to the Initial En Route Qualification training course can be found in Table 4-2 below. None of the competencies can be described as having a “strong” relationship, as defined by Cohen (1977, p. 284).

Table 4-2. Positive Correlations of Competencies of the AT-SAT Test Battery to the Initial En Route Qualification Training Course

	COMPETENCY	DESCRIPTION	R
POSITIVE RELATIONSHIP	ATST (Safety)	Pearson Correlation Sig. (2-tailed) N	.386 .000 137
	Letter Factory (plan)	Pearson Correlation Sig. (2-tailed) N	.344 .000 137
	ATST (Efficiency)	Pearson Correlation Sig. (2-tailed) N	.342 .000 137
	Scanning	Pearson Correlation Sig. (2-tailed) N	.312 .000 137
	Angles	Pearson Correlation Sig. (2-tailed) N	.298 .000 137
	Applied Math	Pearson Correlation Sig. (2-tailed) N	.266 .002 137
	Letter Factory (aware)	Pearson Correlation Sig. (2-tailed) N	.256 .003 137
	Analogies	Pearson Correlation Sig. (2-tailed) N	.254 .003 137
	Dials	Pearson Correlation Sig. (2-tailed) N	.223 .009 137
	ATST (Procedures)	Pearson Correlation Sig. (2-tailed) N	.132 .125 137
	Analogies (Wind)	Pearson Correlation Sig. (2-tailed) N	.081 .348 137
	Analogies (Latency)	Pearson Correlation Sig. (2-tailed) N	.068 .429 137

Hypothesis Three

H_3 There is a negative relationship between the individual subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

The relationship of the individual competencies to the Initial En Route Qualification training course was determined by using bivariate correlation. Eleven competencies were found to have negative relationships. These competencies are listed in Table 4-3 below.

Table 4-3. Negative correlations of Competencies of the AT-SAT Test Battery to the Initial En Route Qualification Training Course

	COMPETENCY	DESCRIPTION	R
NEGATIVE RELATIONSHIP	Experience (Decis)	Pearson Correlation Sig. (2-tailed) N	-.186 .029 137
	Experience (Attent)	Pearson Correlation Sig. (2-tailed) N	-.167 .051 137
	Experience (Thorough)	Pearson Correlation Sig. (2-tailed) N	-.147 .087 137
	Experience (Conf)	Pearson Correlation Sig. (2-tailed) N	-.117 .174 137
	Experience (Cons)	Pearson Correlation Sig. (2-tailed) N	-.114 .183 137
	Experience (Flexi)	Pearson Correlation Sig. (2-tailed) N	-.106 .219 137
	Experience (Comp)	Pearson Correlation Sig. (2-tailed) N	-.087 .312 137
	Experience (Conc)	Pearson Correlation Sig. (2-tailed) N	-.085 .324 137
	Experience (Toler)	Pearson Correlation Sig. (2-tailed) N	-.065 .449 137
	Experience (Execu)	Pearson Correlation Sig. (2-tailed) N	-.048 .579 137
	Experience (Aware)	Pearson Correlation Sig. (2-tailed) N	-.047 .586 137

<u>Null Hypothesis</u>

H_0 There is no clear relationship between the composite score or subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

Having shown that a relationship exists between the composite score on the AT-SAT test battery and the Initial En Route Qualification training course previously in hypotheses one, the question remains if any of the subtests on the AT-SAT test battery show no clear relationship to the Initial En Route Qualification training course. Through the use of multiple regression, the following table describes the variables that were excluded when identifying the most predictive model (Table 4-4).

Table 4-4. Excluded Variables When Utilizing Linear Regression

	Beta In	t	Sig.	Partial Correlation
Dials	.086[c]	.972	.333	.084
Applied Math	.082[c]	.945	.346	.082
Scanning	.113[c]	1.206	.230	.104
LtrFact(aware)	.000[c]	.004	.997	.000
ATST (Eff)	.060[c]	.557	.578	.048
ATST (Proc)	-.011[c]	-.136	.892	-.012
Analogies	.027[c]	.299	.765	.026
Anal (wind)	.053[c]	.680	.498	.059
Anal (Lat)	-.050[c]	-.631	.529	-.055
Exp (Comp)	-.049[c]	-.636	.526	-.055
Exp (Cons)	-.024[c]	-.302	.763	-.026
Exp (Conc)	-.057[c]	-.734	.464	-.064
Exp (Decis)	-.098[c]	-1.253	.213	-.108
Exp (Conf)	-.081[c]	-1.056	.293	-.092
Exp (Toler)	-.023[c]	-.291	.771	-.025
Exp (Execu)	.016[c]	.199	.842	.017
Exp (Thorough)	-.070[c]	-.893	.373	-.078
Exp (Flexi)	-.081[c]	-1.042	.299	-.090
Exp (Aware)	-.002[c]	-.028	.978	-.002
Exp (Attent)	-.107[c]	-1.394	.166	-.120

Summary

This chapter presented the raw data derived from the AT-SAT composite, individual AT-SAT competencies, and Initial En Route Qualification training course scores. Chapter V will present a summary of the study, conclusions, and recommendations associated with this subject.

CHAPTER V

CONCLUSION

Summary

This study was conducted to determine the relationship between the AT-SAT test battery and the scores in the Initial En Route Qualification training course. The relationship between these two measures, as determined by correlation and multiple regression, are indicators of the effectiveness of the AT-SAT test battery as a selection instrument. Numerous external factors, such as motivation, education, intelligence, etc., also influence the success rate of subjects attending the FAA Academy and a difficult or impossible to quantify. These factors were accepted as limitations to this study and the focus has been directed toward the factors that could reasonably be measured and analyzed.

Through the use of correlation and multiple regression, relationships can be better understood and the AT-SAT test battery refined to improve selection by the Federal Aviation Administration. As the AT-SAT test battery and the Initial En Route Qualification training course continue to be utilized, continuous analysis should be implemented. The remaining sections of this chapter will present the conclusions and recommendations derived from this study.

Conclusions

This study was designed to examine the relationship between the AT-SAT test battery and its individual sub-tests to the Initial En Route Qualification training course. The conclusions for each hypothesis will be addressed in the next four sections.

Hypothesis One (H_1) Conclusions

There is a positive relationship between the composite score on the AT-SAT test battery and the composite score in the Initial En Route Qualification training course.

Hypothesis one is accepted because the correlation coefficient of .216 was a positive relationship. The relationship is considered to be moderate; however, the complex nature of the air traffic control profession, and the vast number of external variables that were not included in this study, such as motivation, education, etc., may have an impact on the relationship of the AT-SAT composite score to the Initial En Route Training course score (Cohen, 1977, p. 284). The restriction of range, described in chapter four, has a negative impact on the correlation coefficient. According to Grimm (1993), "Restricted ranges have a tendency to reduce the correlation" (p. 383).

Hypothesis Two (H_2) Conclusions

There is a positive relationship between the individual subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

Hypothesis two is accepted because there were twelve individual AT-SAT competencies with a positive correlation coefficient. As with hypothesis one, none of the relationships can be considered as strong (Cohen, 1977, p. 284). The restriction of range and the other the other external variables, has a negative impact on the correlation coefficient.

Multiple regression was also used to identify a relationship between the Initial En Route Qualification training course scores (dependent variable) and each of the 23 individual AT-SAT battery competencies (independent variables) to determine the strongest model to predict success. The best relationship was attained by correlating the Initial En Route Qualification training course score to the following competencies: Safety, Angles, and Letter Factory (planning). A Pearson's Rho of .469 was achieved at a .05 level of significance. In the behavioral sciences, this is considered a strong relationship, especially considering the restriction of range of the AT-SAT scores (Cohen, 1977, p. 284). The results of this comparison can be seen in Table 5-1 below.

Table 5-1. AT-SAT Competency Models for Strongest Relationship to the Initial En Route Qualification Training Course Scores

MODEL	R	R^2	Std. Error of the Est.
1. ATST(safety)	.386	.149	9.356
2. ATST(safety), Angles	.432	.187	9.180
3. ATST(safety), Angles, Letter Factory(planning)	.469	.220	9.025

The competencies that are excluded when utilizing the multiple regression, model three can be found in Table 5-2 below.

Table 5-2. Excluded Variables when Utilizing Linear Regression

	Beta In	t	Sig.	Partial Correlation
Dials	.086[c]	.972	.333	.084
Applied Math	.082[c]	.945	.346	.082
Scanning	.113[c]	1.206	.230	.104
LtrFact(aware)	.000[c]	.004	.997	.000
ATST (Eff)	.060[c]	.557	.578	.048
ATST (Proc)	-.011[c]	-.136	.892	-.012
Analogies	.027[c]	.299	.765	.026
Anal (wind)	.053[c]	.680	.498	.059
Anal (Lat)	-.050[c]	-.631	.529	-.055
Exp (Comp)	-.049[c]	-.636	.526	-.055
Exp (Cons)	-.024[c]	-.302	.763	-.026
Exp (Conc)	-.057[c]	-.734	.464	-.064
Exp (Decis)	-.098[c]	-1.253	.213	-.108
Exp (Conf)	-.081[c]	-1.056	.293	-.092
Exp (Toler)	-.023[c]	-.291	.771	-.025
Exp (Execu)	.016[c]	.199	.842	.017
Exp (Thorough)	-.070[c]	-.893	.373	-.078
Exp (Flexi)	-.081[c]	-1.042	.299	-.090
Exp (Aware)	-.002[c]	-.028	.978	-.002
Exp (Attent)	-.107[c]	-1.394	.166	-.120

The shared variance (R^2) between the AT-SAT competencies listed as Model 3 above (ATST (safety), Angles, and Letter Factory (planning)) and the Initial En Route Qualification training course score was .220, or approximately 22%. A Venn diagram depicting the common variance between these competencies and the Initial En Route Qualification training course can be found in Figure 5-1 below.

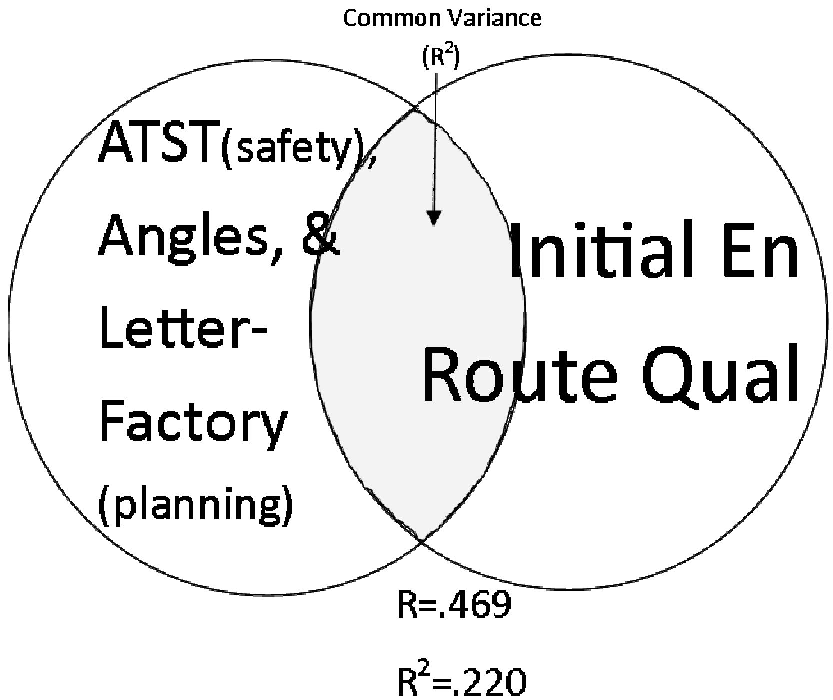

Figure 5-1. Shared Variance (R^2) Between the AT-SAT Competencies (listed as Model 3 above) ATST (safety), Angles, and Letter Factory (planning)) and the Initial En Route Qualification Training Course Score.

<u>Hypothesis Three (H_3) Conclusions</u>

There is a negative relationship between the individual subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

Hypothesis three is accepted because there were eleven individual AT-SAT competencies with a negative correlation coefficient. The individual competencies are: Experience (Decision making), Experience (Attentiveness), Experience (Thoroughness), Experience (Confidence), Experience (Consistency), Experience (Flexibility), Experience

(Composure), Experience (Concentration), Experience (Tolerance), Experience (Execution), and Experience (Awareness).

Null Hypothesis (H_0) Conclusions

There is no clear relationship between the composite score or subtest competencies on the AT-SAT test battery and the composite scores in the Initial En Route Qualification training course.

The null hypothesis is accepted because the AT-SAT competencies Letter Factory (awareness) and Experience (awareness) received correlation coefficients of .000 and .002 respectively when utilizing multiple regression. Two additional AT-SAT competencies, ATST (procedures) and Experience (execution), received correlation coefficients of .012 and .017 respectively, rendering them virtually unrelated to the Initial En Route Qualification training course.

Recommendations

This study was conducted to determine the relationship between the AT-SAT test battery and the Initial En Route Qualification training course. The Initial En Route Qualification training course correlation coefficient of .216 at a.05 level of significance is an indication that the AT-SAT test battery has a positive relationship. However, there are three areas in which improvements can be made in the selection of new air traffic controllers: minimum acceptable score on the AT-SAT test battery, removal/redesign of several individual AT-SAT competencies, and increasing the effectiveness of the AT-

SAT test battery by leveraging the competencies that show the highest relationship to the Initial En Route Qualification training course.

Raise Minimum Acceptable Score for AT-SAT

As described in chapter four, 90.51% (124 out of 137) of the subjects in this study earned a score of 85% or greater, placing them in the "well qualified" category. These candidates achieved an average score that was 5.61% higher in the Initial En Route Qualification training course than the candidates that were placed in the "qualified" category (earning a score 84.99% or less on the AT-SAT test battery). Although additional data is required to confirm, the initial trend shows that a candidate is more likely to pass the Initial En Route Qualification training course when earning a score that places them in the "well qualified" category.

Redesign/Remove Some AT-SAT Test Battery Competencies

This study has shown that several of the individual AT-SAT competencies do not contribute to the effectiveness of the instrument. Most obvious are the competencies related to the experience questionnaire. Of the 23 individual AT-SAT competencies, 11 are derived from the experience questionnaire. The validity of the experience questionnaire is suspect because it is a self-reported evaluation. A candidate's perception of his/her ability cannot be relied upon and many candidates may give answers that they believe are most beneficial to the scoring system instead of a truthful answer. The redesign or removal of some individual AT-SAT competencies has been under consideration since the AT-SAT was integrated into the selection process. Researchers at CAMI continue to test, evaluate, and document the results of new/revised sub-tests on students at the FAA Academy.

Leverage AT-SAT Test Battery Competencies that Show Strong Relationships

Chapter IV described the use of linear regression to identify a relationship between the Initial En Route Qualification training course scores (dependent variable) and each of the 23 individual AT-SAT battery competencies (independent variables). A Pearson's Rho of .469 was achieved at a .05 level of significance by only using competencies Safety, Angles, and Letter Factory (planning). These competencies have high predictive potential and should be used more extensively to select candidates for employment into the air traffic control occupation. Conversely, competencies that showed extremely low predictive power should be eliminated or redesigned.

Future Study

The Initial En Route Qualification training course is the first of four stages of training for an en route air traffic controller. The 137 subjects in this study should be evaluated after August, 2014 to determine the overall effectiveness of the AT-SAT and the Initial En Route Qualification training course. Stages two, three, and four are conducted in field facilities and use a dichotomous scoring system (pass or fail). This longitudinal study would not be able to take advantage of numerical scoring system; however, overall pass rate could be studied.

REFERENCES

Air Traffic Controller Training and Development Group, AJL-11. (2009, November). *Test Analysis for Validation of En Route Initial Training Course 50148001: First Course Conduct*. Washington, DC: Government Printing Office.

Bleckley, K. (2008). *AT-SAT validation against PV.* Copy in possession of the author.

Bleckley, K. (2010, October 5). *2010 AT-CTI "best practices" conference: Air traffic selection and training.* Washington, DC: Government Printing Office.

Booz Allen Hamilton. (2007, June 5). *Learning and development framework air traffic control specialists: Deliverable report.* No. 02EA3TTAGI). Washington, DC: Government Printing Office.

Broach, D. (1996, October 25). *Technical evaluation of the prototype ATSAT selection test battery v1.0 (OAM research task no. AM-97-B-HRR-509).* Washington, DC: Government Printing Office.

Broach, D. (1996, September 17). In Vice L. (Ed.), *ATSAT measurement model workshop.* Copy in possession of the author.

Broach, D. (2009, April 22). *Cut scores in en route air traffic control specialist initial qualifications training.* Washington, DC: Government Printing Office.

Broach, D. (2010, August 27). In Welp R., Kelley R. S.(Eds.), *Evaluation of inter-rater agreement in the en route initial qualifications training graded nonradar labs.* Copy in possession of the author.

Broach, D., & Heil, M. (1997). *Adverse impact and test fairness under the uniform guidelines on employee procedures.* No. AAM-B-HRR-509). Washington, DC: Government Printing Office.

Broach, D., & Manning, C. A. (1997, July). *Review of air traffic controller selection: An international perspective (DOT report no. DOT/FAA/AM-97/15).* Washington, DC: Government Printing Office.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Revised Ed.). New York City, NY: Academic Press.

Dattel, A. R., & King, R. E. (2006, July). *Reweighting AT-SAT to mitigate group score differences.* No. DOT/FAA/AM-06/16). Washington, DC: Government Printing Office.

Department of Labor. (1978a). *29 C.F.R. § 1607.14 technical standards for validity studies. title 29-labor.* Retrieved from http://law.justia.com/cfr/title29/29-4.1.4.1.8.0.22.14.html

Department of Labor. (1978b). *29 C.F.R. § 1607.4 information on impact. title 29-labor.* Retrieved from http://law.justia.com/cfr/title29/29-4.1.4.1.8.0.21.4.html

Department of Transportation - Office of Inspector General. (2008, June 5). *Review of the air traffic controller facility training program (OIG report no. AV-2008-055).* Retrieved from http://www.oig.dot.gov/library-item/5001

Department of Transportation - Office of Inspector General. (2009, June 8). *Training failures among newly hired air traffic controllers (OIG report no. AV-2009-059).* Retrieved from http://www.oig.dot.gov/library-item/4952

Department of Transportation - Office of Inspector General. (2010, April 1). *Review of screening, placement, and initial training of newly hired air traffic controllers (OIG report no. AV-2010-049).* Retrieved from http://www.oig.dot.gov/library-item/5306

Drake, C. A. (1942). *Personnel selection by standard job tests.* New York, NY: McGraw-Hill.

Dunleavy, E. M., Norris, D. G., Baker, D. P., Lentz, E., Horgen, K. E., Tolentino, A. L., & Borman, W. C. (2006, January 25). *ATC tower cab validation study (task order no. DTFAAC-06-T-00004).* Washington, DC: Government Printing Office.

FAA Air Traffic Controller Academy Oversight Group. (2010, December). *Course report for first course conduct of initial en route air traffic training (course 20148001).* Washington, DC: Government Printing Office.

FAA Civil Aeromedical Institute, Human Resources Research Division. (1996a, November). *An alternative measurement and structural model of air traffic control specialist (ATCS) job performance (OAM research task no. AM-97-B-HRR-509).* Washington, DC: Government Printing Office.

FAA Civil Aeromedical Institute, Human Resources Research Division. (1996b, November). *Preliminary evaluation of the prototype ATSAT program test battery v1.0 (OAM research task no. AM-97-B-HRR-509).* Washington, DC: Government Printing Office.

FAA Civil Aeromedical Institute, Human Resources Research Division. (n.d.). *Air traffic selection and training: For the next generation of air traffic control.* Washington, DC: Government Printing Office.

FAA National En Route Redesign Team. (2007, December). *Air traffic training development plan (TDP) for en route initial qualification (stage I) training.* Washington, DC: Government Printing Office.

Federal Aviation Administration. (2010, February 11). *FAA order JO 7110.65T: Air traffic control.* Retrieved from http://www.faa.gov/air_traffic/publications/atpubs/ATC/atc0201.html#atc0201.html.1

Gay, L. R. & Airasian, P. W. (2003). *Educational research: Competencies for analysis and applications* (7th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

Grimm, L. G. (1993, February). *Statistical applications for the behavioral sciences.* New York, NY: Wiley.

Kelley, R. S. (2009, November). *Summary of results of first delivery of en route initial training: Course 50148.* Copy in possession of the author.

Kelley, R. S. (2010a, October). *Course 50148 ERAM lab evaluation grading guidelines.* Copy in possession of the author.

Kelley, R. S. (2010b, October). *Course 50148 nonradar lab evaluation grading guidelines.* Copy in possession of the author.

Kelley, R. S. (2010c, October). *First course conduct 2 (FCC2) evaluation revisions.* Copy in possession of the author.

King, R. E., Manning, C. A. & Drechsler, G. K. (2007, May). *Operational use of the air traffic selection and training battery (report no. DOT/FAA/AM-07/14).* Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA468134

Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels (2nd ed.).* San Francisco, CA: Berrett-Koehler.

Krokos, K. J., Baker, D. P., Norris, D. G., & Smith, M. A. (2007, November 16). *Development of performance standards for air traffic control specialists.* No. 99-G-048). Washington, DC: Government Printing Office.

Kveton, P., Jelinek, M., Vorboril, D., & Klimusova, H. (2007). Computer-based tests: The impact of test design and problem of equivalency. *Computers in Human Behavior,* Retrieved from http://www.sciencedirect.com.ezproxy.osu-tulsa.okstate.edu/science?_ob=ArticleURL&_udi=B6VDC-4C604BH-1&_user=152108&_coverDate=01%2F31%2F2007&_rdoc=1&_fmt=full&_orig=search&_origin=search&_cdi=5979&_sort=d&_docanchor=&view=c&_acct=C000012538&_version=1&_urlVersion=0&_userid=152108&md5=82e52a7a119a2911e5ff4e23bea59b08&searchtype=a&artImgPref=F

Mitchell, M. (4/6/2010). *FAA hiring practices for air traffic controllers in question.* Retrieved from http://avstop.com/news_april_2010/faa_hiring_practices_for_air_traffic_controllers_in_question.htm

Nickels, B. J., Bobko, P., Blair, M. D., Sands, W. A., & Tartak, E. L. (1995, January). *Separation and control hiring assessment (SACHA) final job analysis report (contract no. DTFA01-C-00032).* Washington, DC: Government Printing Office.

Popham, W. J. (1992). *Educational evaluation (3rd ed.).* Needham Hieghts, MA: Allyn-Bacon.

Popham, W. J. (2008, April). *Transformative assessment*. Alexandria, VA: ASCD.

Porter, D., Cole, D. R., Barton, M., & Kieckhaefer, W. (1996, April 8). *The letter factory test: A proposed measure of situational awareness, planning, thinking ahead, problem identification, prioritization, and recall from interruption, version 2.0.* Washington, DC: Government Printing Office.

Quinnipiac (n.d.). *Pearson's r correlation – A rule of thumb*. Retrieved from http://faculty.quinnipiac.edu/libarts/polsci/Statistics.html

Ramos, R. A., Heil, M. C., & Manning, C. A. (2001a, March). *Documentation of validity for the AT-SAT computerized test battery volume I (DOT report no. DOT/FAA/AM-01/5).* Washington, DC: Government Printing Office.

Ramos, R. A., Heil, M. C., & Manning, C. A. (2001b, March). *Documentation of validity for the AT-SAT computerized test battery volume II (DOT report no. DOT/FAA/AM-01/5).* Washington, DC: Government Printing Office.

Schmitt, N., Borman, W. C., & Associates. (1993). *Personnel selection in organizations*. San Francisco, CA: Jossey-Bass.

Sethumadhavan, A., & Durso, F. T. (2009, February). Selection in air traffic control: Is nonradar training a predictor of radar performance? Retrieved from http://search.ebscohost.com.ezproxy.osu-tulsa.okstate.edu/login.aspx?direct=true&db=iih&AN=40101447&site=ehost-live

Smith, M. & Robertson, I. T. (1993). *The theory and practice of systematic personnel selection (2nd ed.).* London: Macmillan Press LTD.

Triola, M. F. (2007). *Elementary statistics using EXCEL* (3rd ed.). Boston, MA: Pearson Education.

Tsacoumis, S., Anderson, L. E., & King, R. E. (2006, September). *Final report: Development of an equivalent alternate form of the air traffic selection and training battery.* Washington, DC: Government Printing Office.

U. S. General Accounting Office. (1991). *FAA staffing: Better strategy needed to ensure facilities are properly staffed..* (Testimony before the Subcommittee on Government Activities and Transportation, Committee on Government Operations, House of Representatives No. GAO/T-RCED-92-8). Washington, DC: Government Printing Office.

U. S. General Accounting Office. (1997). *Aviation safety: Opportunities exist for FAA to refine the controller staffing process.* (Report to Congressional Requesters No. GAO/RCED-97-84). Washington, DC: Government Printing Office.

U. S. General Accounting Office. (2008). *Federal aviation administration: Efforts to hire, staff, and train air traffic controllers are generally on track, but challenges remain.* (Testimony before the Subcommittee on Aviation, Committee on Transportation and infrastructure, House of Representatives No. GAO-08-908T). Washington DC: Government Printing Office.

Wanous, J. P. (1992). *Recruitment, selection, orientation, and socialization of newcomers (2nd ed.).* USA: Addison-Wesley.

Welp, R., Broach, D., & Kelley, R. S. (2010, December). *Test development and analysis for validation of initial en route air traffic training (course 50148001) first course conduct.* Washington, DC: Government Printing Office.

Welp, R., Kelley, R. S., Doskow, J., Rounsavell, L., & Morrison, M. (2007, November). *En route initial qualification (stage I) training redesign front-end analysis: Requirements, deficiencies, and opportunities for improvement.* Washington, DC: Government Printing Office.

Wiersma, W., (1995). *Research methods in education: An introduction* (6th ed.). Needham Heights, MA: Allyn & Bacon.

Wiggins, M. W., & Stevens, C. (1999). *Aviation social science: Research methods in practice*. Brookfield, VA: Ashgate.

APPENDICES

APPENDIX A

RESEARCH AGREEMENTS

A-1: RESEARCH AGREEMENT WITH THE FEDERAL AVIATION ADMINISTRATION’S EXECUTIVE DIRECTOR FOR HUMAN RESOURCES PROGRAMS AND POLICIES

A-2: RESEARCH AGREEMENT WITH THE CIVIL AEROSPACE MEDICAL INSTITUTE (CAMI)

A-3: RESEARCH AGREEMENT WITH THE OKLAHOMA STATE UNIVERSITY INSTITUTIONAL REVIEW BOARD

APPENDIX A-1. RESEARCH AGREEMENT WITH THE FEDERAL AVIATION ADMINISTRATION'S EXECUTIVE DIRECTOR FOR HUMAN RESOURCES PROGRAMS AND POLICIES

Mr. Kelley,

Thanks for your patience. After consultation with AGC, I am able to give you the go-ahead for your study, with some caveats.

First, the FAA cannot give you any race, national origin, sex/gender/sexual orientation, age or other protected-class information about any of the people who took AT-SAT. Our intent is that you will not conduct analyses of how AT-SAT relates to these factors.

Second, we request an advance copy of your study, especially well prior to publication.

Please allow me to introduce Dr. Lexsee Waterford, who is responsible for testing matters for the Office of Human Resource Management. I am copying her on this e-mail. Please contact her in future on this subject. I would appreciate it if you would copy me though, since I was involved initially and would like to continue to follow your progress.

Best wishes on your dissertation,

Jay Aul
Acting Executive Director for
Human Resources Programs and Policies
Federal Aviation Administration
(202) 267-9862
Jay.Aul@faa.gov

APPENDIX A-2. RESEARCH AGREEMENT WITH THE CIVIL AEROSPACE MEDICAL INSTITUTE (CAMI)

Research Agreement

A key research problem in aviation safety is the selection of personnel into safety-critical occupations such as air traffic control specialist (ATCS). Analysis of the data from different perspectives helps to advance our scientific knowledge and understanding of this applied problem. To that end and to advance scientific understanding in personnel selection, de-identified ATCS selection data are requested for analysis by Ronald Scott Kelley under the supervision of Todd Hubbard, Ed.D., Dissertation Chair, Oklahoma State University.

Ronald Scott Kelley will ensure that the requested data will be used for the stated research purpose only. The data will be protected from inadvertent disclosure, nor it will they be re-distributed to any other party. Ronald Scott Kelley will provide a copy of the Oklahoma State University Institutional Review Board finding to the FAA. All presentations, publications, briefings and other artifacts created in the course of the research will include an acknowledgment that the data were provided by the Federal Aviation Administration's Civil Aerospace Medical Institute. The acknowledgment will also include a disclaimer that the opinions expressed are those of the author alone and do not necessarily reflect the official policies of the United States government. Ronald Scott Kelley will provide a bound copy of the completed and approved dissertation to the Civil Aerospace Medical Institute Library.

Agreed

For the FAA

M. Kathryn Bleckley, Ph.D.

Printed Name

Signature

6 June 2011

Date

For the User

Ronald Scott Kelley

Printed Name

Signature

June 7, 2011

Date

Doctoral Dissertation Supervisor

Todd Hubbard, Ed.D.,
Oklahoma State University

Printed Name

Signature

June 7, 2011

Date

APPENDIX A-3. RESEARCH AGREEMENT WITH THE OKLAHOMA STATE UNIVERSITY INSTITUTIONAL REVIEW BOARD

REC'D URC
JUN 08 2011

Oklahoma State University Institutional Review Board

Request for Determination of Non-Human Subject or Non-Research

Federal regulations and OSU policy require IRB review of all research involving human subjects. Some categories of research are difficult to discern as to whether they qualify as human subject research. Therefore, the IRB has established policies and procedures to assist in this determination.

1. **Principal Investigator Information**

First Name: Ronald	Middle Initial: Scott	Last Name: Kelley
Department/Division: Applied Educational Studies (Aviation and Space Science)		College: Graduate
Campus Address: N/A		Zip+4: N/A
Campus Phone: N/A	Fax: N/A	Email: rscottkelley@yahoo.com
Complete if PI does not have campus address:		
Address: 453 Suzan Road		City: Tuttle
State: Oklahoma	Zip: 73089	Phone: 405-517-6864

2. **Faculty Advisor** **(complete if PI is a student, resident, or fellow)** ☐ **N/A**

Faculty Advisor's name: Todd Hubbard, Ed.D.	Title: Dissertation Chair, ASO Professor
Department/Division: Applied Educational Studies (Aviation and Space Science)	College: Graduate
Campus Address: 318 Willard Hall Stillwater, OK	Zip+4: 74078
Campus Phone: (405)744-8062 Fax: N/A	Email: Todd.Hubbard@okstate.edu

3. **Study Information:**

A. Title

Relationship between Air Traffic Selection and Training (AT-SAT) battery test scores and success in the Initial En Route Air Traffic Control Qualification training course at the Federal Aviation Administration (FAA) Academy.

B. Give a brief summary of the project. (See instructions for guidance)

This study will describe the relationship between the AT-SAT test battery, a predictive test used prior to selection of Air Traffic Control candidates, and the success in the Initial En Route Air Traffic Control Qualification training course, a course designed to teach and measure actual Air Traffic Control job functions.

The three research hypotheses are:

H_1 There is a strong relationship between the composite score on the AT-SAT test battery and the success in the Initial En Route Air Traffic Control Qualification training course.

H_2 There is a strong relationship between many of the subtests on the AT-SAT test battery and the success in the Initial En Route Air Traffic Control Qualification training course.

H_3 There is a weak relationship between many of the subtests on the AT-SAT test battery and the success in the Initial En Route Air Traffic Control Qualification training course.

Revision Date: 04/2006 3 of 5

Student data will be provided by the FAA's Civil Aerospace Medical Institute (CAMI) to the PI. All data will be coded to protect the identity of all subjects (see section C). The hypotheses will be tested through the use of multiple regression.

C. Describe the subject population/type of data/specimens to be studied. (See instructions for guidance)

The subject population consists of Air Traffic Control Students that completed the Initial En Route Air Traffic Control Qualification Training at the Federal Aviation Administration Academy. Approximately 100 students will be used in this study. The students age range from 19 to 31 years of age.

The identity of the students will be kept confidential by the FAA's Civil Aerospace Medical Institute (CAMI). CAMI will provide the PI a coded identity (Student 1, Student 2, etc.) of the test subjects and the requested data associated with each student. CAMI will not provide any information that may allow any link to the identification of students.

CAMI will provide the following data from each student: 1) All test scores from the Air Traffic Selection and Training battery (AT-SAT), and 2) All test scores from the Initial En Route Air Traffic Control Qualification Training course.

4. Determination of "Research".
45 CFR 46.102(d): *Research* means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. Activities which meet this definition constitute research for purposes of this policy whether or not they are conducted or supported under a program which is considered research for other purposes.

One of the following must be "no" to qualify as "non-research":

A. Will the data/specimen(s) be obtained in a systematic manner?
☐ No ☒ Yes

B. Will the intent of the data/specimen collection be for the purpose of contributing to generalizable knowledge (the results (or conclusions) of the activity are intended to be extended beyond a single individual or an internal program, e.g., publications or presentations)?
☐ No ☒ Yes

5. Determination of "Human Subject".
45 CFR 46.102(f): *Human subject* means a living individual about whom an investigator (whether professional or student) conducting research obtains: (1) data through intervention or interaction with the individual or (2) identifiable private information. Intervention includes both physical procedures by which data are gathered (for example venipuncture) and manipulations of the subject or the subject's environment that are performed for research purposes. Interaction includes communication or interpersonal contact between investigator and subject. Private information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be individually identifiable (i.e., the identity of the subject is or may be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects.

A. Does the research involve obtaining information about living individuals?
☐ No ☒ Yes
If no, then research does not involve human subjects, no other information is required.
If yes, proceed to the following questions.

All of the following must be "no" to qualify as "non-human subject":

B. Does the study involve intervention or interaction with a "human subject"?

Oklahoma State University Institutional Review Board

Request for Determination of Non-Human Subject or Non-Research

☒ No ☐ Yes

C. Does the study involve access to identifiable private information?
☒ No ☐ Yes

D. Are data/specimens received by the Investigator with identifiable private information?
☒ No ☐ Yes

E. Are the data/specimen(s) coded such that a link exists that could allow the data/specimen(s) to be re-identified?
☐ No ☒ Yes
If "*Yes*," is there a written agreement that prohibits the PI and his/her staff access to the link?
☐ No ☒ Yes

6. Signatures

Signature of PI [signature] **Date** 6-7-11

Signature of Faculty Advisor [signature] **Date** 6-7-11
(If PI is a student)

☒ Based on the information provided, the OSU-Stillwater IRB has determined that this project **does not** qualify as human subject research as defined in 45 CFR 46.102(d) and (f) and **is not subject to oversight by the OSU IRB.**

☐ Based on the information provided, the OSU-Stillwater IRB has determined that this research **does** qualify as human subject research and **submission of an application for review by the IRB is required.**

[signature] 6-9-11
Dr. Shelia Kennison, IRB Chair Date

Revision Date: 04/2006 5 of 5

APPENDIX B

DANA BROACH, PH.D., ANALYSIS OF INTER-RATER RELIABILITY ON THE NEW GRADING INSTRUMENT

Federal Aviation Administration

Memorandum

Date:	August 27, 2010
To:	Scott Kelley, AJL-11
	Robert Welp, Ph.D., AJL-11
From:	Dana Broach, Ph.D., AAM-520
Subject:	Evaluation of Inter-rater Agreement in the En Route Initial Qualifications Training Graded Non radar Labs

Purpose

The purpose of this memorandum is to document the degree of inter-rater agreement attained in the evaluation of student performance in the en route non-radar graded labs conducted on August 12, 2010 at the FAA Academy. A new observable behavior evaluation method was used by two independent raters to evaluate student performance. Details on the evaluation method are available from the lead for the en route redesign.

Method

There were seven (7) students in the en route class. Two graded non-radar labs were administered to each student; each student was observed and evaluated by two independent raters.

Data available for analysis. A percent score was generated from each evaluator. In addition to the observed performance, each evaluator independent assessed the student's performance qualitatively. The qualitative categories were Fail (Badly), Fail (Marginal), Pass (Marginal), and Pass (Good).

Statistical analysis. Two analyses were performed. First, Rater 1 and Rater 2 percent scores were correlated across students and labs. A high degree of correlation suggests good inter rater agreement; a low correlation suggests poor inter-rater agreement. Second, Cohen's kappa was estimated across rater pairs using the SPSS CROSSTAB procedure. Generally, a Cohen's kappa greater than .7 is considered as an acceptable level of inter-rater agreement [LeBreton, J. M. & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815-852. (DOI: 10.1177/1094428106296642)].

Results

The overall Pearson correlation between Rater 1 and Rater 2 scores across students and labs was .885 (p<.001; n=14). Scores given by Rater 1 are plotted against those given by Rater 2 in Figure 1 in the attachment. The absolute largest difference in scores was 6.47 percentage points.

The crosstab of qualitative judgments are presented in Table 1 in the attachment. Out of the 14 instances (7 students by 2 labs), raters disagreed in only 2 instances. In the first disagreement, one rater predicted a "Pass Marginal" and the other predicted "Fail Marginal." In the second, one rater predicted a "Pass Marginal" and the other predicted a "Pass Good." Cohen's kappa was .736 (p<.001; n=14). LeBreton and Senter characterize this level of inter-rater agreement as "strong."

Conclusion

The new evaluation method for evaluating and reporting student performance in the graded non-radar labs in the en route initial qualifications training course demonstrates sufficient inter-rater agreement for use with future classes with job jeopardy attached. Continued evaluation of inter-rater agreement over classes is highly recommended.

Dana Broach, Ph.D.
Personnel Research Psychologist

Attachment (Figure & Table)

Figure 1: Scatterplot of Rater 1 score by Rater 2 score across students

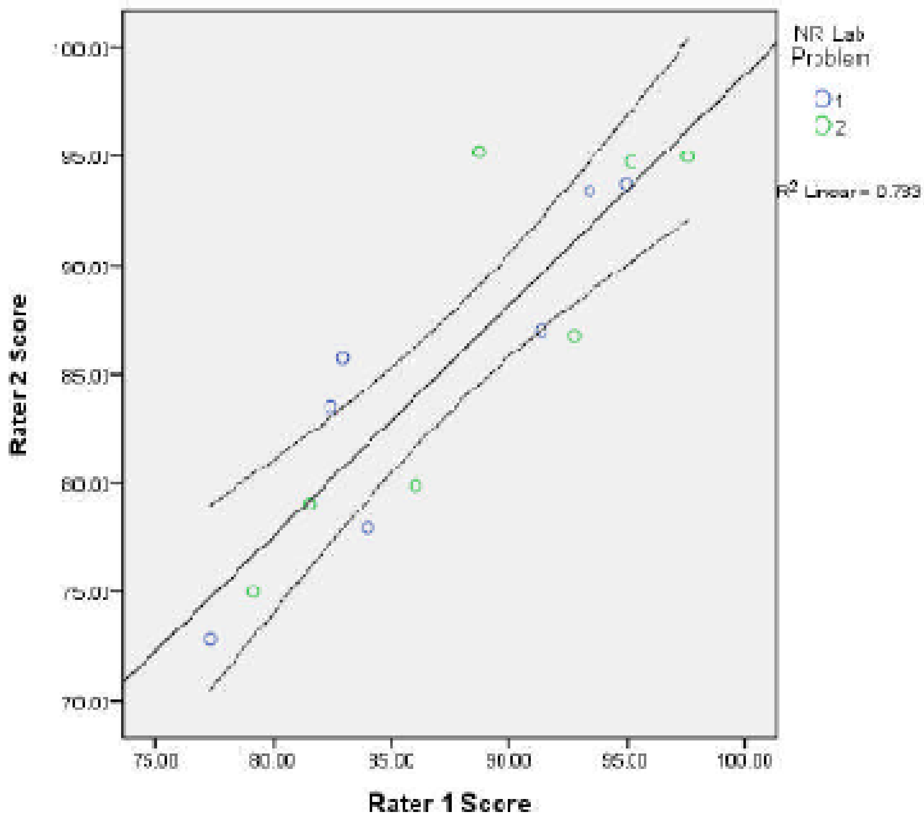

Table 1: Cross-tabulation of Rater 1 by Rater 2 predicted student performance across students and non-radar labs

		Rater 1 Predicted Student Performance			
		Fail Marginal	Pass Marginal	Pass Good	Total
Rater 2 Predicted Student Performance	Fail Marginal	1	1	0	2
	Pass Marginal	0	3	1	4
	Pass Good	0	0	8	8
	Total	1	4	9	14

APPENDIX C

DEVELOPMENT AND VALIDATION

OF AT-SAT

Because the process was so complex, a summary of the development and validation will be described beginning with a six step process described by Ramos, Heil, and Manning (2001a) in the Documentation of validity for the AT-SAT computerized test battery volume I and II.

Step 1: Complete predictor battery development. Through the use of the Separation and Control Hiring Assessment (SACHA) project and other related documents, several prototype predictor tests were develop for the worker requirements that were deemed the most important. A team of air traffic control subject matter experts selected 12 tests and developed these into a test battery (Alpha). The decision was made to limit the validation effort to the en route option because the validation could occur on a compressed schedule and the sample size could be managed more effectively (Ramos, Heil, & Manning, 2001a, p. 2).

Step 2: Complete criterion measure development. Three job measures (supervisor and peer ratings of typical performance, computerized job sample, and performance on a high-fidelity simulation of the ATCS job) were developed to evaluate en route job performance. The high-fidelity simulator was used to evaluate construct validity because it was the most realistic environment over the other two criterion measures (Ramos, Heil, & Manning, 2001a, p. 3).

Step 3: Conduct concurrent validation study. A high positive correlation was demonstrated between AT-SAT test scores and the job performance of a large sample of en route air traffic controllers. The study was continued and expanded to include all the en route facilities so the required data could be attained (Ramos, Heil, & Manning, 2001a, p. 4).

Step 4: Conduct pseudo-applicant study. Since air traffic controllers are a select group, it was likely that the range of scores produced by current air traffic controllers would be restricted. By having predictor scores that have a low degree of variability, it is expected that the mean scores would be higher than a random sample from the general population. The true validity of the selection battery is likely to underestimate the correlation to the job performance measures. To offset the results from the restricted (air traffic controllers) sample, validity coefficients were used to more closely reflect what the real benefits of the test battery would be for an unrestricted (general public) applicant population. Military and civilian pseudo-applicants were utilized to correct the initial estimates for validity and to obtain some initial estimates of race and/or gender bias (Ramos, Heil, & Manning, 2001a, p. 4).

Step 5: Analysis and validation of predictors. Data management was considered a critical element in the validation process. The goal of the team was to complete the analysis within two weeks of the final data collection. The methods used by the team to determine the validity of the predictors and decide the final test composite were predictor-criterion relationships and reviews of the individual elements. The final AT-SAT battery composition included the sub-tests with the highest correlation to job performance and the least variability between protected classes (Ramos, Heil, & Manning, 2001a, p. 4).

Step 6: Deliver predictor battery and supporting documentation. According to Ramos, Heil, and Manning (2001a), "The goal of developing a selection test battery for the ATCS that was highly job related and fair to women and minorities was achieved" (p. 5).

The Alpha Battery was pilot tested with 14 sub-tests. These sub-tests were: Air Traffic Scenarios test, Sound test, Letter Factory test, Dials test, Static Vector/Continuous Memory test, Experiences Questionnaire, Time Wall/Pattern Recognition test, Analogy test, Classification test, Word Memory test, Scan test, Planes test, Angles test, and Applied Mathematics test. Some of the sub-test did not survive the validation process. The sub-tests in the Alpha battery will be discussed in further detail in the following paragraphs (Ramos, Heil, & Manning, 2001a, p. 23).

As stated in the previous paragraph, not all sub-tests in the Alpha battery were successfully validated. According to Ramos, Heil, and Manning (2001a), each sub-test was evaluated and was considered for elimination if they exhibited any of the following characteristics:

> ***Low Discrimination:*** The item did not discriminate between those individuals who received high versus low total scores, stated as a biserial correlation. ***Check Option:*** One or more incorrect response options had positive biserial correlations with total test score. ***Too Hard:*** The percent correct was low. ***Too Easy:*** The percent correct was high. ***High Omits:*** The item was skipped or not reached, with these two problems being distinguishable from each other (p. 37).

The following table (Table C-1) describes the results of the analysis for the Alpha battery of the AT-SAT (Ramos, Heil, & Manning, 2001a, pp. 37-54).

Table C-1. Results of the Analysis for the Alpha Battery of the AT-SAT

SUB-TEST	DESCRIPTION	RESULTS
Applied math sub-test.	Applicants taking less than five seconds scored extremely low and approximately half scored below chance.	Test cut from 53 to 30 items and re-ordered.
Dials sub-test.	Applicants taking less than 4.25 seconds per item were not taking enough time to read the questions and/or not putting forth their best effort.	Test cut from 57 to 44 items.
Angles sub-test.	The item analysis did not reveal any problem items and there appeared to be a good distribution of item difficulties.	No edits.
Sound sub-test.	The item analysis did not reveal any problem items and there appeared to be a good distribution of item difficulties.	An alternative scoring procedure, based on the number of within-item digits correct with partial credit for digit reversals, was recommended for the beta version.
Memory sub-test.	The item analysis did not reveal any problem items and there appeared to be a good distribution of item difficulties.	No edits.
Analogy sub-test.	Word items. Since the analogy items based on the number of syllables performed poorly, this type of item was not used when replacing the non-semantic word items. Instead, the five non-semantic word items were replaced with *combinations of specific letters* and *phonetic* items. Additionally, three semantic items were replaced with three new semantic items of more reasonable (expected) difficulty levels. Visual items. Since the non-semantic picture items demonstrated a relatively stable alpha (.67) and high item-total correlations, no items were removed. In an effort to stabilize the alpha further, three non-semantic picture items were added, increasing the non-semantic visual subtest from 13 to 16 items.	Six non-semantic word items were removed due to low item-total correlations, five being *syllable* items (i.e., the correct solution to the analogy was based on number of syllables). Seven more items were removed from the alpha Analogy test version due to either very high or low difficulty level, or to having poor distracter items.

Table C-1 (Continued)

Classification sub-test.	Of the original 46 items, only three of the four scales (i.e., Semantic Word, Semantic Visual, and Non-Semantic Visual) and a total of 22 items contributed sufficiently to test reliability to warrant inclusion in a revised test version. To construct a test having the same three parts and increase the reliability to about .80 (for number-correct scores), the number of items would need to increase from the 22 to 139. It was further found that the Classification test correlates highly with the Analogy test.	Given that the Classification test had lower reliability scores than the Analogy test, it was recommended that the Classification test be eliminated from the AT-SAT battery.
Letter factory sub-test.	The original plan was to measure three worker requirements using the LFT. Because the measure of Recall From Interruption showed ceiling effects and unreliable difference scores, it was recommended that attempts to measure that worker requirement with this test be abandoned.	To more adequately measure the worker requirements of Planning and Thinking Ahead and Situational Awareness, lengthening the test to 93 minutes was recommended. This longer version includes doubling the number of practice sequences that participants complete before they begin the test. It was estimated that this extra practice would reduce the practice effect observed between the LFT and the retest LFT on a small (N = 184) subsample. This would help ensure that participants perform at or near their ability prior to beginning the test portion of the LFT.
Scan sub-test.	Items in the Scan test that change during their screen presentation did not behave the same as other items in the test. Eliminating those items improved estimates of internal consistency reliability. After eliminating four items that had poor item-total correlations, the 162 remaining items in the test (i.e., non-practice) portion of the Scan test produced an alpha of .96.	It was recommend to keep the Scan test at its current length and allocating 21 minutes for test completion.

Table C-1 (Continued)		
Planes sub-test.	The project team cut the number of items in each part of the original Planes test in half for the alpha data collection effort. This was done to meet project time constraints. After completing reliability analyses, it was clear that the test would benefit from restoring it to its original length.	The number of items in Part 3 and in the practice sessions was cut in half. The time allotted for breaks between the three test parts was also halved.
Experience questionnaire.	The EQ results in the pilot test were promising. Most of the scales looked good in terms of their means, variances, and reliabilities. The two scales that were weakest, psychometrically, were Self-Awareness and Self-Monitoring/Evaluating.	Item analysis suggested that items 21, 53, and 163 should be deleted, and item 144 moved to a different scale.
Air traffic scenarios sub-test.	It was felt if separate scores were to be used in the concurrent validation, additional practice and test trials would be needed to achieve a high level of reliability for the "Separation Skill" variable.	It was recommended that three practice trials be used with each trial targeted to test understanding of specific rules and more tailored feedback after each trial.
Time wall / pattern recognition sub-test.	The three scores analyzed for the TW test were (a) Pattern Recognition Accuracy (PRACCY), defined as the percent of correct pattern matching responses out of all correct and incorrect responses (e.g., excluding time-outs); (b) Pattern Recognition Speed (PRSPD), a trans-formation of the average time, in milliseconds, for correct responses; and (c) Time Wall Accuracy (TWACCY), a transformation of the mean absolute time error, in milliseconds.	Time Wall Accuracy reliability estimates were modest, although the test-retest correlations held up fairly well. Preliminary results suggested that five or six trials may be needed to get highly reliable results on all three measures.
Static Vector/Continuous Memory sub-test.	No data could be located in the literature.	The sub-test appears to be eliminated from the test battery.

To validate the AT-SAT test scores against actual job performance, current air traffic controllers were evaluated on several measures. Although many criterion scores were collected, the final measures used were: the Computer-Based Performance Measure (CBPM), the Behavior Summary Scales, and the High Fidelity Performance Measure (HiFi). The HiFi consisted of two independent scores. A brief description of each can be found below in Table C-2, Criterion for ATC job performance (Ramos, Heil, & Manning, 2001b, p. 37).

Table C-2. Criterion for ATC Job Performance

SCORING INSTRUMENT	DESCRIPTION
Computer-Based Performance Measure (CBPM)	A medium fidelity simulation. A computer displayed a simulated air space sector while the examinee answered questions based on the air traffic scenario shown.
Behavior Summary Scales	Performance ratings completed by the examinee's peers and supervisors.
High Fidelity Performance Measure – 1 (HiFi)	Observers' comprehensive ratings of the examinee's two-day performance on a high-fidelity air traffic control simulator (Core Technical score - a composite of several scores).
High Fidelity Performance Measure – 2 (HiFi)	Observers' comprehensive ratings of the examinee's two-day performance on a high-fidelity air traffic control simulator (Controlling Traffic Safely and Efficiently - a composite of several scores).

According to Ramos, Heil, and Manning (2001b):

> The small sample size for the HiFi measures precluded their use in the selection of a final predictor battery and computation of the predictor composite. They were used, however, in some of the final validity analyses as a comparison standard for the other criteria. A single, composite criterion was computed using the CBPM score and the composite Ratings score. Thus, the following three criteria were used for the validity analyses: (a) the CBPM score, (b) the composite Ratings score, and (c) the composite criterion score (p.37).

APPENDIX D

DEVELOPMENT AND VALIDATION OF THE INITIAL EN ROUTE QUALIFICATION COURSE

The first delivery of the course was conducted from August 14 to November 5, 2009, at the Mike Monroney Aeronautical Center, Oklahoma City, OK. A testing and evaluation strategy was developed that split the course in two distinct parts. A pass/fail point was added at the end of part one, meaning that, students that did not achieved a score of 70 percent or greater would be removed from the program and their employment with the FAA would be terminated. An overview of the testing and evaluation process, with weighting can be seen in Table D-1, Scored events during the first delivery (FAA Air Traffic Controller Academy Oversight Group, 2010, p. 6).

Table D-1. Scored Events During the First Delivery

Event	Description	% of score
Part 1 – Non Radar		
Map Test	Tests knowledge of FAA Academy airspace	10
Controller Knowledge Test 1	Tests knowledge of the lecture lessons taught covering fundamental en route air traffic control concepts and procedures	15
Three Nonradar Lab Evaluation Problems	Three application tests in which students must apply air traffic concepts and procedures to nonradar scenarios	75
Part 2 –Radar		
Controller Knowledge Test 2	Tests knowledge of the lecture lessons taught covering radar related en route air traffic control concepts and procedures	15
Skill Checks 1 and 2	Two separate events that assess the student's ability to make basic En Route Automation Modernization (ERAM) entries	10
Three Radar Lab Evaluation Problems	Three application tests in which students, performing in the Radar Associate position, must apply air traffic concepts and procedures to radar scenarios	75

As part of the FAA's Instructional System Design (ISD) process, this first delivery, called an operational tryout, is used to document to ensure the course validity (i.e., technically correct and effective) prior to regular delivery. Normally, the students that participate in an operational tryout are not given credit for a course; however, given the length of this course, it was not feasible to follow this process. The students were given credit for the course; but, since the course had not yet been validated, no students were failed or removed from employment (i.e., under a job jeopardy condition) (Kelley, 2009, p. 1). Every aspect of the course was evaluated. According to the FAA Air Traffic Controller Academy Oversight Group (2010) responsible for the course, the method for data collection was:

- All lesson material was reviewed by several air traffic control specialists for ***completeness and accuracy***. All discrepancies or omissions were noted for correction.
- Student ratings and comments (see sample form below) were collected after every lecture lesson, eLearning module, and lab exercise for the ***effectiveness*** of the lesson and ***student acceptance*** (includes such measures as quality of materials, pace of instruction, interest level, etc.).
- End-of Lesson tests were analyzed to determine if test items were valid. While all test items were discussed with students, any test item that was missed by 30% or more of the students (5 or more students) was reviewed to determine why such a result occurred. Inaccurate or poorly worded test questions were either re-worded and re-tested or deleted.
- Student scores on scored events were recorded to determine the ***effectiveness*** of the instruction.
- Nonradar and radar lab instructors were interviewed in a group setting to obtain their observations and recommendations regarding the ***effectiveness*** of the labs.
- The Academy collected data on ghost pilot and remote pilot operator errors and ERAM operational problems during the radar labs.
- The En Route Redesign Team reviewed student test scores and lab evaluation scenarios to determine the ***adequacy*** of the course (p. 5).

Overall, the course was well received by the students. A new element introduced into the course was the use of eLearning exercises. These exercises were used to compliment the instructor led lessons. The students and instructors gave the eLearning exercises high marks. The students and lab instructors did not give the nonradar lab good ratings. The students reported not receiving sufficient practice to adequately prepare for the nonradar evaluations. The grading instrument used also exhibited problems. The major issue was the fact that students were given 100 points at the beginning of the scenario and points were deducted for each error committed. In many cases, the students had numbers that were negative (below zero), indicating the overall possible score was above 100 points. The grading instrument used for the radar portion of the course was adequate and passed validation (Kelley, 2009, pp. 7-8).

Multiple adjustments were made prior to the second delivery of the course. The mid-course exit point was removed. The testing and evaluation events (see Table 3-4) was re-weighted to have a single exit point at the end of the course to make the pass or fail determination. The nonradar scenarios were reduced from three to two graded scenarios. An aircraft characteristics test was added; and the two skills checks were combined into one (Kelley, 2009, pp. 7-8).

The most significant difference in the course between the first and second delivery was in the nonradar labs. The evaluation instrument was redesigned to allow a flexible scoring system. Each student was given credit for each item completed correctly or points deducted for each item performed incorrectly and a ratio of correct to incorrect items was calculated. This system more accurately reflects the performance of the student. Nine additional instructional scenarios were added to address the lack of practice. The weighted scores and events that were used in the second delivery of the course are provided below in Table D-2 (FAA Air Traffic Controller Academy Oversight Group, 2010, pp. 13-14).

Table D-2. Revised Weighted Scores for Second Delivery

REVISED WEIGHTED SCORES FOR SECOND DELIVERY		
Event	**Description**	**% of score**
Map Test	Tests knowledge of FAA Academy airspace	3%
Controller Knowledge Test 1	Tests knowledge of the lecture lessons taught covering fundamental en route air traffic control concepts and procedures	5%
Two Nonradar Lab Evaluation Problems	Two application tests in which students must apply air traffic concepts and procedures to nonradar scenarios	14% (7% each)
Aircraft Characteristics Test	Tests knowledge of aircraft characteristics	5%
Controller Knowledge Test 2 -	Tests knowledge of the lecture lessons taught covering radar related en route air traffic control concepts and procedures	6%
Computer Skills Checklist	Assess the student’s ability to make basic En Route Automation Modernization (ERAM) entries	7%
Three Radar Lab Evaluation Problems	Three application tests in which students, performing in the Radar Associate position, must apply air traffic concepts and procedures to radar scenarios	60% (20% each)

The second delivery of the course was conducted from June 29 to September 24, 2010, at the Mike Monroney Aeronautical Center, Oklahoma City, OK. The high number of edits to technical content identified during the first delivery, coupled with the new scoring system prevented this course from being conducted as a pass/fail course. Consequently, as with the first delivery, the students were not in job jeopardy. The same data collection process was used as in the first delivery, with the exception of the radar associate labs because it had passed validation (Kelley, 2010, p. 1).

The course scored very high ratings from the students and instructors. There were very few edits from the course; however, the redesign team acknowledged that the students had very high scores going into the last two days of the course (the radar associate evaluations). The final three evaluations are the capstone of the course. The students demonstrate nearly all prior skills learned during these evaluations. The weighted scores that were developed as the final weights for the course prior to delivering the course on a regular basis can be found below in Table D-3, Final weighted scores (FAA Air Traffic Controller Academy Oversight Group, 2010, pp. 19-22).

Table D-3.   Final Weighted Scores

FINAL WEIGHTED SCORES		
Event	**Description**	**% of score**
Map Test	Tests knowledge of FAA Academy airspace	2%
Controller Knowledge Test 1	Tests knowledge of the lecture lessons taught covering fundamental en route air traffic control concepts and procedures	4%
Two Nonradar Lab Evaluation Problems	Two application tests in which students must apply air traffic concepts and procedures to nonradar scenarios	14% (7% each)
Aircraft Characteristics Test	Tests knowledge of aircraft characteristics	4%
Controller Knowledge Test 2	Tests knowledge of the lecture lessons taught covering radar related en route air traffic control concepts and procedures	5%
Computer Skills Checklist	Assess the student's ability to make basic En Route Automation Modernization (ERAM) entries	5%
Three Radar Lab Evaluation Problems	Three application tests in which students, performing in the Radar Associate position, must apply air traffic concepts and procedures to radar scenarios	66% (22% each)

Validation. Validation efforts were conducted throughout the design, development, and delivery of the course. According to the FAA Air Traffic Controller Academy Oversight Group (2010):

> Each test was subjected to strict validation requirements to ensure the course met the standards required by the Uniform Guidelines on Employee Selection, 29 CFR 1607. A consultant from the Civil Aerospace Medical Institute (CAMI) participated throughout the test development process to provide guidance and clarification of these requirements (p. 4).

The steps taken during the development of the knowledge tests and performance evaluations will be explained in greater detail.

Knowledge tests. As each lesson was developed, a team of subject matter experts developed a list of questions to be considered for each end-of-lesson test. The questions were reviewed by instructional design specialists. The questions were revised, if necessary, then accepted or deleted. All test items are linked to instructional objectives to ensure that each test question is related to the job duties required for air traffic controllers. A test blueprint was developed to select the type and number of test items to be used on the Block tests and the comprehensive Controller Knowledge Tests (Welp, Broach, & Kelley, 2010, p. 5).

Test item analysis was conducted during the first and second deliveries of the course. Instructors documented the number of correct responses to each question, and followed up on any items missed by 30 percent or more of the students. According to Welp, Broach, and Kelley (2010):

> The purpose of the discussion was to determine if the question was misleading or incorrect, the instruction was unclear or missing, or if some combination of these two problems existed. In addition, students were asked to evaluate the adequacy and fairness of the test and provide comments during the discussion. Corrective action for problematic questions was then identified (pp. 5-6).

Problematic questions were re-written and re-tested later in the course. Instruction judged unclear or inadequate was revised.

During the second delivery of the course, two additional versions of each test were developed to allow a random sample of test questions. Students were administered the second and third versions of the test questions and the same test item analysis technique was followed as in the first delivery of the course (Welp, Broach, & Kelley, 2010, pp. 5-6).

The knowledge tests were found to be valid. As documented by the Air Traffic Controller Training and Development Group, AJL-11 (2009):

> At the conclusion of the second delivery of the En route Initial Training course, there was no indication that the test questions are problematic. This is supported by:
>
> - No test questions were missed by more than 29% of the students.
> - Instructors did not identify any erroneous or problematic questions.
>   *Note: During discussion, students identified one test question in Lesson 17 as having more than one possible correct answer. Instructors agreed that, while no students missed the question and no student scores were negatively affected, the question should be revised and rechecked.*
> - Student evaluations of the end-of-lesson tests, block tests, and overall assessment of testing for the second delivery of the course were unanimous in indicating the tests are both adequate and fair (p.7).

Performance evaluations. There were two performance evaluations developed for the course: one for performance in the nonradar labs and one for performance in the radar associate labs. Each instrument used for the evaluation is unique and was designed independently; therefore, they will be described separately (Welp, Broach, & Kelley, 2010, pp. 10-11).

The nonradar evaluations were conducted at the end of the nonradar section of the course. During the first delivery of the course, three nonradar evaluations were conducted. Following first delivery, two evaluations were determined to be sufficient and a change was made prior to the beginning of the second delivery of the course (Welp, Broach, & Kelley, 2010, p. 11).

FAA Form 3120-25 is the evaluation instrument used at operational facilities to guide and document training performance. An example of the form can be found in Figure D-1 below.

ATCT/ARTCC OJT
INSTRUCTION/EVALUATION REPORT

1. Name	2. Date	3. Scenario/Position(s)

4. Weather
- ☐ VFR
- ☐ MVFR
- ☐ IFR
- ☐ Other___________

5. Workload
- ☐ Light
- ☐ Moderate
- ☐ Heavy

6. Complexity
- ☐ Not Difficult
- ☐ Occasionally Difficult
- ☐ Mostly Difficult
- ☐ Very Difficult

7. Hours

8. Total Hours This Position

9. Purpose ☐ OJT ☐ OJF ☐ Familiarization Scenario ☐ Instructional Scenario ☐ Evaluation Scenario
☐ Skill Check ☐ Certification ☐ Recertification ☐ Skill Enhancement ☐ Other

10. Routing

11. Performance

Job Task	Job Subtask	Observed	Comment	Satisfactory	Needs Improvement	Unsatisfactory	Simulation Training
A. Separation	1. Separation is ensured.						
	2. Safety alerts are provided.						
B. Coordination	3. Performs handoffs/pointouts.						
	4. Required coordinations are performed.						
C. Control Judgment	5. Good control judgment is applied.						
	6. Priority of duties is understood.						
	7. Positive control is provided.						
	8. Effective traffic flow is maintained.						
D. Methods and Procedures	9. Aircraft identity is maintained.						
	10. Strip posting is complete/correct.						
	11. Clearance delivery is complete/correct and timely.						
	12. LOAs/directives are adhered to.						
	13. Additional services are provided.						
	14. Rapidly recovers from equipment failures and emergencies.						
	15. Scans entire control environment.						
	16. Effective working speed is maintained.						
E. Equipment	17. Equipment status information is maintained.						
	18. Equipment capabilities are utilized/understood.						
F. Communication	19. Functions effectively as a radar/tower team member.						
	20. Communication is clear and concise.						
	21. Uses prescribed phraseology.						
	22. Makes only necessary transmissions.						
	23. Uses appropriate communications method.						
	24. Relief briefings are complete and accurate.						
G. Other							

FAA Form 3120-25 (9-96) Supersedes Previous Edition

Figure D-1. FAA Form 3120-25.

The redesign team developed a modified FAA Form 3120-25 to quantify the errors committed by the students during their training and evaluations. According to Welp, Broach, and Kelley (2010):

> Subjective items were eliminated to ensure the objective nature of the evaluation. For example, working speed was NOT a graded item because the errors generated from not performing in time were graded. Students were given 100 points at the beginning of the scenario and points were deducted each time a student made an error (pp. 7-10).

The nonradar evaluation form used for the first delivery can be found in Figure D-2, Nonradar grading form (first delivery), below.

U.S. Department of Transportation
Federal Aviation Administration

50148 - NON RADAR

INSTRUCTION/EVALUATION REPORT

1. Name	2. Date	3. Scenario/Position(s)

4. Weather	5. Workload	6. Complexity	7. Hours
☒ MVFR	☒ Moderate	☒ Occasionally Difficult	0+30

9. Purpose ☐ Instructional ☒ ◻ Evaluation	10. Class #

	Job Task	Job Subtask	Comment	EVALUATION Errors	Points	Total (Errors X Points)
Performance	A. Separation	1-a.Separation is ensured – Aircraft to aircraft or MEA violations.			12	
		1-b.Separation is ensured – Aircraft to airspace.			10	
	B. Coordination	4. Required coordinations are performed.			8	
	C. Control Judgment	8. Effective traffic flow is maintained.			7	
	D. Methods and Procedures	9. Aircraft identity is maintained.			6	
		10. Strip posting is complete/correct.			2	
		11. Clearance delivery is complete/correct and timely.			4	
		12. LOAs/directives are adhered to.			4	
	F. Communication	20. Communication is clear and concise.			2	
		21. Uses prescribed phraseology.			2	

12. Comments *COMMENT SECTION NOT TO SCALE*

Instructor/Evaluator Signature: ______________________________ Date:

Student Signature: ______________________________. Date:

Figure D-2. Non Radar Grading Form (first delivery).

Prior to using the nonradar grading instrument on “real” students, several practice sessions were conducted for the instructors/evaluators using instructors as “mock students”. On each practice run, two instructors observed the same performance. The “mock students” made errors intentionally to give the evaluators practice using the instrument and to resolve scoring differences. This process was repeated until “the differences noted by the evaluators were found to be within acceptable variance” (Welp, Broach, and Kelley, 2010, p. 11).

The results of the first delivery of the course indicated some problems existed with the nonradar scoring instrument. Two evaluators were used during the grading (this time using actual students attending the course) and the differences noted were within acceptable range; however, the scoring system was not found to be adequate by the redesign team. Welp, Broach, and Kelley (2010) added:

> The new grading instrument developed for the nonradar labs was considered by many (students and observers) to be “overly harsh” and did not adequately reflect students’ capability, knowledge, and skill to apply nonradar procedures. The average score for the nonradar evaluations was 38.46%. The lowest score was 0% (It was possible to lose more than 100 points and negative scores were possible and did occur. Negative scores, when they did occur, were rounded up to “0”) and the highest score was 96%. The evaluators were also asked to rate each students performance based on the following assessment scale: PASS(good), PASS(marginal), FAIL(marginal), or FAIL(poor). This assessment was used to determine if the numerical scores determined by the evaluation instrument matched the opinions of subject matter experts. The instrument and the expert assessment did NOT agree and further identified the need to redesign the evaluation instrument (p. 11).

A new scoring instrument was developed that allowed individual scoring based on the actions of each student. Welp, Broach, and Kelley (2010) described the instrument:

> Each sub-task performed by the student is graded. During the scenario, if a student performs a sub-task correctly, positive points are assessed. If the student performs a sub-task incorrectly, negative points are accessed (this includes the

> failure to take an action, such as a delay). At the end of the evaluation, all positive points are added and divided by the sum of all positive and negative points. This score is represented as a percentage and points are assessed as this percentage. For example, if a student performed 90 correct actions and 21 incorrect actions, the student would receive a score of 81.08% (90/111) (pp. 11-12).

The revised nonradar grading form used during second delivery of the course can be found in Figure D-3, Nonradar grading form (second delivery) below:

COURSE 50148 NONRADAR LAB SCENARIO # _____ STUDENT NAME ____________________ DATE ________

EVALUATOR NAME (PRINT) ______________________ EVALUATOR SIGNATURE ______________________

ARRIVALS

INITIAL CHECK IN							
1. Acknowledge Check On							
2. WX/ATIS/Altimeter							
INBOUND COORDINATION							
3. Interphone Usage (format and initials)							
4. Call Sign/Type/Suffix							
5. Estimate							
6. Altitude and Restrictions Forwarded							
7. Approach Type							
8. Other [D67 (block & cancel), HKS, Etc.]							
9. TCP							
ARRIVAL CLEARANCE							
10. Clearance Limit							
11. Route							
12. Altitude **AIRSPACE** Separation, Enter ++ or --							
13. Fix/Mileage **AIRSPACE** Separation, Enter ++ or --							
14. Altitude **AIRCRAFT** Separation, Enter ++ or --							
15. Fix/Mileage **AIRCRAFT** Separation, Enter ++ or --							
HOLDING (16-19)							
16. Assigned Altitude							
17. Direction (and Turns)							
18. Radial/Bearing or "As Published"							
19. EFC							
20. Frequency Change							
21. Efficiency (Correct Rule/No Delay)							
APPROACH CLEARANCE							
22. Altitude **AIRSPACE** Separation, Enter ++ or --							
23. Fix/Mileage **AIRSPACE** Separation, Enter ++ or --							
24. Altitude **AIRCRAFT** Separation, Enter ++ or --							
25. Fix/Mileage **AIRCRAFT** Separation, Enter ++ or --							
26. Approach Name							
27. Frequency Change							
28. Efficiency (Correct Rule/No Delay)							
OTHER							
29. Stripmarking							
30. Phraseology (Including NE vs. NW)							
31. Board Management							
TOTAL +							=____(1)
TOTAL –							
TOTAL + AND –							=____(2)

NOTES

DEPARTURES

COORDINATION							
32. EDCT	___	___	___	___	___	___	
33. Interphone Usage	___	___	___	___	___	___	
34. Call Sign (Including "In Suspense")	___	___	___	___	___	___	
35. Airport/Departure Time/Estimate	___	___	___	___	___	___	
36. Altitude	___	___	___	___	___	___	
37. Revised Routing	___	___	___	___	___	___	
DEPARTURE CLEARANCE							
38. Takeoff/Turns Confirmation, **Enter ++ or --**	___	___	___	___	___	___	
39. Departure/Destination Airport Name	___	___	___	___	___	___	
40. Initial Departure Instructions	___	___	___	___	___	___	
41. Route	___	___	___	___	___	___	
42. Altitude **AIRSPACE Separation, Enter ++ or --**	___	___	___	___	___	___	
43. Fix/Mileage **AIRSPACE Separation, Enter ++ or --**							
44. Altitude **AIRCRAFT Separation Enter ++ or --**	___	___	___	___	___	___	
45. Fix/Mileage **AIRCRAFT Separation Enter ++ or --**	___	___	___	___	___	___	
46. Within 2000' of Req. Alt. & not IAFDOF	___	___	___	___	___	___	
47. Void Time	___	___	___	___	___	___	
48. "Verify This Clearance…" **Enter ++ or --**	___	___	___	___	___	___	
49. Frequency Change	___	___	___	___	___	___	
50. Efficiency (Correct Rule/No Delay)	___	___	___	___	___	___	
OTHER							
51. Stripmarking	___	___	___	___	___	___	
52. Phraseology (Including NE vs. NW)	___	___	___	___	___	___	
53. Board Management	___	___	___	___	___	___	

ALL OTHER AIRCRAFT

OTHER							
54. Interphone Usage	___	___	___	___	___	___	
55. Acknowledge Check On	___	___	___	___	___	___	
56. WX/ATIS/Altimeter	___	___	___	___	___	___	
57. Coordination	___	___	___	___	___	___	
58. Revised Routing (If Any)	___	___	___	___	___	___	
59. Altitude **AIRSPACE Separation, Enter ++ or --**	___	___	___	___	___	___	
60. Fix/Mileage **AIRSPACE Separation, Enter ++ or --**	___	___	___	___	___	___	
61. Altitude **AIRCRAFT Separation, Enter ++ or --**	___	___	___	___	___	___	
62. Fix/Mileage **AIRCRAFT Separation, Enter ++ or –**	___	___	___	___	___	___	
63. IAFDOF	___	___	___	___	___	___	
64. Frequency Change	___	___	___	___	___	___	
65. Efficiency (Correct Rule/No Delay)							
66. Stripmarking	___	___	___	___	___	___	
67. Phraseology (Including NE vs. NW)	___	___	___	___	___	___	
68. Board Management	___	___	___	___	___	___	
TOTAL +							= (3)
TOTAL –	___	___	___	___	___	___	
TOTAL + AND –	___	___	___	___	___	___	=___(4)

Line (1) + Line (3) = ________ (5)
Line (2) + Line (4) = (6)
Line (5) ÷ Line (6) = ________ %

Figure D-3. Non Radar Grading Form (second delivery).

Practice sessions were conducted to test the feasibility of the grading instrument. The results of the sessions showed the instrument to be more complex but after practice, the instructors were able to become proficient and were favorable to the changes. Two evaluators were used for the grading of the students during the second delivery of the course. As with the first delivery, the instructors compared scores and the differences were within acceptable variance. As documented by Welp, Broach, and Kelley (2010):

> The average score for the nonradar evaluation was 86.68%. The lowest score was 75.07% and the high score was 96.28%. The two main reasons for the improved evaluation score when compared to the first delivery are: 1) the students were given nine additional instructional scenarios, and 2) the grading instrument measured positive and negative actions instead of only deducting points from an artificially low total of all possible actions. The evaluators were also asked to rate each students performance based on the following assessment scale: PASS(good), PASS(marginal), FAIL(marginal), or FAIL(poor), This assessment was used to determine if the numerical scores determined by the evaluation instrument matched the opinions of subject matter experts. The instrument and the expert assessment strongly agree on 26 out of 28 individual assessments, while the remaining two assessments were only marginally skewed (p. 14).

A personnel research psychologist from the Civil Aerospace Medical Institute (CAMI), conducted an independent analysis of inter-rater reliability on the new grading instrument (Appendix B). According to Broach (2010):

> Method
> There were seven (7) students in the en route class. Two graded nonradar labs were administered to each student; each student was observed and evaluated by two independent raters.
>
> Data available for analysis.
> A percent score was generated from each evaluator. In addition to the observed performance, each evaluator independent assessed the student's performance qualitatively. The qualitative categories were Fail (Badly), Fail (Marginal), Pass (Marginal), and Pass (Good).
>
> Statistical analysis.
> Two analyses were performed. First, Rater 1 and Rater 2 percent scores were correlated across students and labs. A high degree of correlation suggests good

inter-rater agreement; a low correlation suggests poor inter-rater agreement. Second, Cohen's kappa was estimated across rater pairs using the SPSS CROSSTAB procedure. Generally, a Cohen's kappa greater than .7 is considered as an acceptable level of inter-rater agreement [LeBreton, J. M.& Senter, J. L. (2008). Answers to 20 questions about inter-rater reliability and inter-rater agreement. *Organizational Research Methods*, *11*(4), 815-852. (DOI: 10.1177/1094428106296642)].

Results
The overall Pearson correlation between Rater 1 and Rater 2 scores across students and labs was .885 ($p<.001$; $n=14$). Scores given by Rater 1 are plotted against those given by Rater 2 in Figure 1 in the attachment. The absolute largest difference in scores was 6.47 percentage points.

The crosstab of qualitative judgments are presented in Table 1 in the attachment. Out of the 14 instances (7 students by 2 labs), raters disagreed in only 2 instances. In the first disagreement, one rater predicted a "Pass Marginal" and the other predicted "Fail Marginal." In the second, one rater predicted a "Pass Marginal" and the other predicted a "Pass Good." Cohen's kappa was .736 ($p<.001$; $n=14$). LeBreton and Senter characterize this level of inter-rater agreement as "strong."

Conclusion
The new evaluation method for evaluating and reporting student performance in the graded nonradar labs in the en route initial qualifications training course demonstrates sufficient inter-rater agreement for use with future classes with job jeopardy attached. Continued evaluation of inter-rater agreement over classes is highly recommended (pp. 1-2).

An example of a completed nonradar grading form used during second delivery of the course can be found in Figure D-4, Completed Nonradar grading form (second delivery) below:

COURSE 50148 NONRADAR LAB SCENARIO # 1 STUDENT NAME Joe Student DATE 10-3-11

EVALUATOR NAME (PRINT) Frank Evaluator EVALUATOR SIGNATURE Frank Evaluator

ARRIVALS	N211K	AAL12	N52KL				
INITIAL CHECK IN							
1. Acknowledge Check On	+	+	−				
2. WX/ATIS/Altimeter	+	+	−				
INBOUND COORDINATION							
3. Interphone Usage (format and initials)	+	+	+				
4. Call Sign/Type/Suffix	−	+	+				
5. Estimate	+	+	+				
6. Altitude and Restrictions Forwarded	+	−	+				
7. Approach Type	+	−	+				
8. Other [D67 (block & cancel), HKS, Etc.]							
9. TCP	+	+	+				
ARRIVAL CLEARANCE							
10. Clearance Limit		+					
11. Route		+					
12. Altitude **AIRSPACE** Separation, Enter ++ or --		− −					
13. Fix/Mileage **AIRSPACE** Separation, Enter ++ or --		++					
14. Altitude **AIRCRAFT** Separation, Enter ++ or --		++					
15. Fix/Mileage **AIRCRAFT** Separation, Enter ++ or --		++					
HOLDING (16-19)							
16. Assigned Altitude							
17. Direction (and Turns)							
18. Radial/Bearing or "As Published"							
19. EFC							
20. Frequency Change							
21. Efficiency (Correct Rule/No Delay)							
APPROACH CLEARANCE							
22. Altitude **AIRSPACE** Separation, Enter ++ or --	++		++				
23. Fix/Mileage **AIRSPACE** Separation, Enter ++ or --	++		++				
24. Altitude **AIRCRAFT** Separation, Enter ++ or --	++		++				
25. Fix/Mileage **AIRCRAFT** Separation, Enter ++ or --	++		++				
26. Approach Name	−		+				
27. Frequency Change	+		−				
28. Efficiency (Correct Rule/No Delay)	+		+				
OTHER							
29. Stripmarking	−	+	+				
30. Phraseology (Including NE vs. NW)	−	+	+				
31. Board Management	+	+	+				
TOTAL +	18	17	19				= 54 (1)
TOTAL −	4	4	3				
TOTAL + AND −	22	21	22				= 65 (2)

DEPARTURES

	BUFF01	ASE2011				
COORDINATION						
32. EDCT	+	+				
33. Interphone Usage	+	+				
34. Call Sign (Including "In Suspense")	+	+				
35. Airport/Departure Time/Estimate	+	–				
36. Altitude	+	–				
37. Revised Routing	+	–				
DEPARTURE CLEARANCE						
38. Takeoff/Turns Confirmation, **Enter ++ or --**	++	++				
39. Departure/Destination Airport Name	–	+				
40. Initial Departure Instructions	–	+				
41. Route	+	+				
42. Altitude **AIRSPACE** **Separation, Enter ++ or --**	++	--				
43. Fix/Mileage **AIRSPACE** **Separation, Enter ++** or **--**	++	++				
44. Altitude **AIRCRAFT** **Separation Enter ++ or --**	++	++				
45. Fix/Mileage **AIRCRAFT** **Separation Enter ++** or --	++	--				
46. Within 2000' of Req. Alt. & not IAFDOF	–	+				
47. Void Time	+	+				
48. "Verify This Clearance…" **Enter ++ or --**	++	++				
49. Frequency Change	+	+				
50. Efficiency (Correct Rule/No Delay)	+	+				
OTHER						
51. Stripmarking	+	+				
52. Phraseology (Including NE vs. NW)	+	+				
53. Board Management	+	+				

ALL OTHER AIRCRAFT

	A51967	DAL57	N510TQ				
OTHER							
54. Interphone Usage	+	+	+				
55. Acknowledge Check On	+	+	+				
56. WX/ATIS/Altimeter	+	+	+				
57. Coordination		–					
58. Revised Routing (If Any)		–					
59. Altitude **AIRSPACE** **Separation, Enter ++ or --**		++					
60. Fix/Mileage **AIRSPACE** **Separation, Enter ++** or --		++					
61. Altitude **AIRCRAFT** **Separation, Enter ++ or --**		++					
62. Fix/Mileage **AIRCRAFT** **Separation, Enter ++** or --		++					
63. IAFDOF							
64. Frequency Change	+	+	+				
65. Efficiency (Correct Rule/No Delay)		+					
66. Stripmarking	+	+	+				
67. Phraseology (Including NE vs. NW)		+					
68. Board Management	+	+	+				
TOTAL +	31	37	6				= 74 (3)
TOTAL –	3	9	0				
TOTAL + AND –	34	46	6				= 86 (4)

Line (1) + Line (3) = 128 (5)
Line (2) + Line (4) = 151 (6)
Line (5) ÷ Line (6) = 84.77 %

Figure D-4. Completed Nonradar Grading Form (second delivery).

The radar associate evaluations were conducted at the end of the course (over a two day period). As with the nonradar instrument, a modified FAA Form 3120-25 was developed to quantify the errors committed by each student and subjective items were deleted. Two evaluators practiced using the new instrument with instructors serving as "mock students" as in the nonradar practice. Welp, Broach, and Kelley (2010) concluded, "The differences noted by the evaluators were found to be within acceptable variance" (p. 16).

The grading of the students during the first delivery of the course yielded the following results as described by Welp, Broach, and Kelley (2010):

> Two evaluators were used during the actual grading of the students during the first delivery of the course. The two evaluators graded the scenario and compared their final scores and the errors noted on the grade forms. Although the differences noted by the evaluators were found to be within acceptable variance, the evaluators noted some differences in the assessment of errors. After discussion between the evaluators, the differences were reconciled. The observed mistakes were consistent between evaluators; however, the category of error that was documented was at times different. It was determined that the grading guidelines needed further explanation to eliminate future problems. The average score for the radar associate evaluations was 73.13%. The lowest score was 0% and the highest score was 97%. The evaluators were also asked to rate each students performance based on the following assessment scale: PASS(good), PASS(marginal), FAIL(marginal), or FAIL(poor). This assessment was used to determine if the numerical scores determined by the evaluation instrument matched the opinions of subject matter experts. The expert assessments agree on 54 out of 56 individual assessments, while the remaining 2 assessments were only marginally skewed. The rater reliability for the evaluation instrument was within acceptable variance. It should be noted, the rater reliability outliers were generally found on scenarios that students performed very poorly (pp. 16-17).

An example of the modified FAA Form 3120-25 can be found below in Figure D-5, Radar associate grading form.

U.S. Department of Transportation
Federal Aviation Administration

50148 – RADAR ASSOCIATE INSTRUCTION/EVALUATION REPORT

1. Name	2. Date	3. Scenario/Position(s)	
4. Weather ☒ MVFR	5. Workload ☒ Moderate	6. Complexity ☒ Occasionally Difficult	7. Hours 0+45
9. Purpose ☐ Instructional ☒ Evaluation			10. Class #

	Job Task	Job Subtask	Comment	EVALUATION Error	Points	Total (Errors X Points)
Performance	A. Separation	1-a. Separation is ensured – Aircraft to aircraft or MEA violations.			16	
		1-b. Separation is ensured – Aircraft to airspace.			12	
	B. Coordination	3. Performs handoffs/pointouts.			9	
		4. Required coordinations are performed.			8	
	C. Control Judgment	5. Good control judgment is applied.			5	
		6. Priority of duties is understood			5	
		7. Positive control is provided.			5	
		8. Effective traffic flow is maintained.			4	
	D. Methods and Procedures	9. Aircraft identity is maintained.			6	
		10. Strip posting is complete/correct.			2	
		11. Clearance delivery is complete/correct and timely.			4	
		12. LOAs/directives are adhered to.			4	
		13. Additional services are provided.			3	
		14. Rapidly recovers from equipment failures / emergencies			2	
		15. Scans entire control environment.			2	
		16. Effective working speed is maintained.			N/A	
	E. Equipment	17. Equipment status information is maintained.			2	
		18. Equipment capabilities are utilized/understood.			2	
	F. Communication	19. Functions effectively as a radar/tower team member.			2	
		20. Communication is clear and concise.			2	
		21. Uses prescribed phraseology.			2	
		22. Makes only necessary transmissions.			2	
		23. Uses appropriate communications method.			2	
		24. Relief briefings are complete and accurate.			5	
	G. Other					

Figure D-5. Radar Associate Grading Form.

There were no major revisions to the radar associate grading form; however, further guidance and instructions were developed to aid evaluators in determining the type of errors that should be noted on the instrument. The second delivery of the course was conducted using the same instrument and no additional concerns were identified. An example of a completed form can be found below in Figure D-6, Completed radar associate grading form.

U.S. Department of Transportation
Federal Aviation Administration

50148 – RADAR ASSOCIATE INSTRUCTION/EVALUATION REPORT

1. Name	2. Date	3. Scenario/Position(s)
Joe Student	11-4-11	RA #2

4. Weather	5. Workload	6. Complexity	7. Hours
☒ MVFR	☒ Moderate	☒ Occasionally Difficult	0+45

9. Purpose ☐ Instructional ☒ Evaluation

10. Class # 12345

	Job Task	Job Subtask	Comment	EVALUATION: Error	EVALUATION: Points	EVALUATION: Total (Errors X Points)
Performance	A. Separation	1-a. Separation is ensured – Aircraft to aircraft or MEA violations.		1	16	16
		1-b. Separation is ensured – Aircraft to airspace.		Ø	12	
	B. Coordination	3. Performs handoffs/pointouts.		Ø	9	
		4. Required coordinations are performed.		Ø	8	
	C. Control Judgment	5. Good control judgment is applied.		Ø	5	
		6. Priority of duties is understood		Ø	5	
		7. Positive control is provided.		Ø	5	
		8. Effective traffic flow is maintained.		Ø	4	
	D. Methods and Procedures	9. Aircraft identity is maintained.		Ø	6	
		10. Strip posting is complete/correct.		3	2	6
		11. Clearance delivery is complete/correct and timely.		Ø	4	
		12. LOAs/directives are adhered to.		Ø	4	
		13. Additional services are provided.		Ø	3	
		14. Rapidly recovers from equipment failures / emergencies		Ø	2	
		15. Scans entire control environment.		1	2	2
		16. Effective working speed is maintained.			N/A	
	E. Equipment	17. Equipment status information is maintained.		Ø	2	
		18. Equipment capabilities are utilized/understood.		Ø	2	
	F. Communication	19. Functions effectively as a radar/tower team member.		Ø	2	
		20. Communication is clear and concise.		Ø	2	
		21. Uses prescribed phraseology.		1	2	2
		22. Makes only necessary transmissions.		Ø	2	
		23. Uses appropriate communications method.		Ø	2	
		24. Relief briefings are complete and accurate.		Ø	5	
	G. Other					-26 = 74%

Figure D-6. Completed Radar Associate Grading Form.

After the conclusion of the second delivery, the FAA Air Traffic Controller Academy Oversight Group (2010) wrote, “the results of the evaluation indicate the course has met all validation requirements” (p. 1).

VITA

Ronald Scott Kelley

Candidate for the Degree of

Doctor of Education

Thesis: RELATIONSHIP BETWEEN AIR TRAFFIC SELECTION AND TRAINING (AT-SAT) BATTERY TEST SCORES AND COMPOSITE SCORES IN THE INITIAL EN ROUTE QUALIFICATION TRAINING COURSE AT THE FEDERAL AVIATION ADMINISTRATION (FAA) ACADEMY

Major Field: Applied Educational Studies

Biographical:

Education: Completed the requirements for the Bachelor of Science in Business Administration at East Central Oklahoma State University, Ada, OK, 1987; completed the requirements for the Master of Science in Aerospace Administration at Southeastern Oklahoma State University, Durant, OK in 2007; completed the requirements for the Doctor of Education in Applied Educational Studies at Oklahoma State University, Stillwater, Oklahoma in May, 2012.

Experience: Federal Aviation Administration, March, 1998 to present. Supervisory Air Traffic Control Specialist at Memphis ARTCC and the FAA Academy, Air Traffic Control Training Specialist at Washington Headquarters, FAA Academy Instructor at the FAA Academy, and Air Traffic Control Specialist at Houston ARTCC, Denver ARTCC, and Fort Worth ARTCC, Adjunct Instructor, Southeastern Oklahoma State University, August, 2008 to March, 2011.

Professional Memberships: Federal Aviation Administration Managers Association.

Name: Ronald Scott Kelley Date of Degree: May, 2012

Institution: Oklahoma State University Stillwater, Oklahoma

Title of Study: RELATIONSHIP BETWEEN AIR TRAFFIC SELECTION AND TRAINING (AT-SAT) BATTERY TEST SCORES AND COMPOSITE SCORES IN THE INITIAL EN ROUTE QUALIFICATION TRAINING COURSE AT THE FEDERAL AVIATION ADMINISTRATION (FAA) ACADEMY

Pages in Study: 130 Candidate for the Degree of Doctor of Education

Major Field: Applied Educational Studies

Scope and Method of Study: This study focused on the development and use of the AT-SAT test battery and the Initial En Route Qualification training course for the selection, training, and evaluation of air traffic controller candidates. The Pearson product moment correlation coefficient was used to measure the linear relationship between the AT-SAT composite scores and the composite scores of the Initial En Route Qualification Training course. Multiple regression was used to determine the relationship between the Initial En Route Qualification Training course and the AT-SAT test battery sub-tests.

Findings and Conclusions: The correlation coefficient between the AT-SAT test battery composite score and the Initial En Route Qualification training course was .216. There were twelve individual AT-SAT sub-tests with positive relationships to the Initial En Route Qualification training course. Eleven sub-test competencies were found to have a negative relationship to the Initial En Route Qualification training course and two competencies had no clear relationship.

ADVISOR’S APPROVAL: Dr. Todd Hubbard