EMBRY-RIDDLE
Aeronautical University™
SCHOLARLY COMMONS

Theses - Daytona Beach   Dissertations and Theses

12-18-1993

Comparison of Air Traffic Control Candidate Ability with Simulator-Based Training Measures

Lawrence A. Tomaskovic

Embry-Riddle Aeronautical University - Daytona Beach

Follow this and additional works at: https://commons.erau.edu/db-theses

Part of the Multi-Vehicle Systems and Air Traffic Control Commons

---

COMPARISON OF AIR TRAFFIC CONTROL CANDIDATE

ABILITY WITH SIMULATOR-BASED TRAINING MEASURES

by

Lawrence A. Tomaskovic

A Thesis Submitted to the
Office of Graduate Programs
in Partial Fulfillment of the Requirements for the Degree of
Master of Aeronautical Science

Embry-Riddle Aeronautical University
Daytona Beach, Florida
December 18, 1993

UMI Number: EP31863

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.
In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

COMPARISON OF AIR TRAFFIC CONTROL CANDIDATE ABILITY WITH SIMULATOR-BASED TRAINING MEASURES

by

Lawrence A. Tomaskovic

This thesis was prepared under the direction of the candidate's thesis committee chairman, Dr. Gerald D. Gibb, Center for Aviation/Aerospace Research, and has been approved by the members of his thesis committee. It was submitted to the Office of Graduate Programs and was accepted in partial fulfillment of the requirements for the degree of Master of Aeronautical Science.

THESIS COMMITTEE:

Dr.Gerald D. Gibb,
Chairman

Dr.John A. Wise,
Member

Mr. Maryin L. Smith,
Member

Department Chair, Aeronautical Science

Dean of Faculty, Daytona Beach campus   1-21-94

                                                        Date

Acknowledgements

I wish to extend my heartfelt thanks to my thesis chairman, Dr. Gerald Gibb, for his unselfish and unending assistance with my thesis. Even with his busy schedule, he always had time to answer a question or to provide an explanation. I would also like to thank my thesis committee member, Dr. John Wise. He opened up the world of Human Factors to me which allowed new insights and answers to certain problems that I only had a vague knowledge. I would also like to extend my heartfelt thanks to Mr. Marvin Smith, who encouraged me to continue my graduate studies, supported me along the way and introduced me to the academic profession. I would also like to thank my co-workers and friends, Messrs. Esa Rantanen and Kenneth Petschauer. Esa and Ken helped to shoulder the burden of setting up the ATC lab and then helped with teaching the classes and keeping the lab running "properly". The long hours we put in forge the bonds that make for a strong friendship. Finally, a special thanks to my father, Mr. Andrew Tomaskovic. It was his support, that only a father gives to a son, that kept me focused on my education.

Abstract

Author: Lawrence A. Tomaskovic

Title: Comparison of Air Traffic Controller Candidate Ability

               With Simulator-Based Training Measures
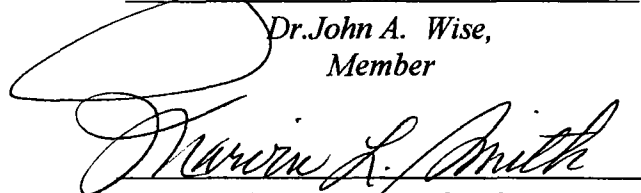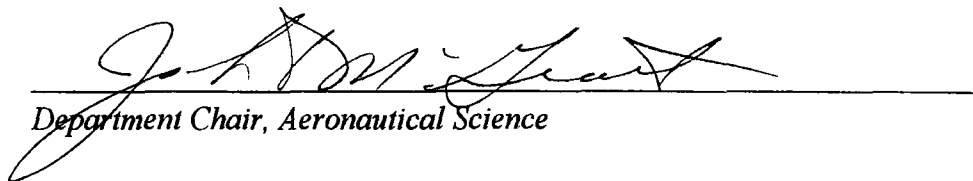
Institution: Embry-Riddle Aeronautical University

Degree: Master of Aeronautical Science

Year: 1993

The purpose of this study was to determine if the utilization of an experimental computer-based selection test battery would aid in the prediction of a candidates performance when using an air traffic control computer-based simulation program. Each candidate completed the selection test battery, and then received air traffic control instruction using the air traffic control simulation program incorporated in the TRACON/Pro™ simulator system. The selection test battery results were correlated with the subsequent simulator scoring results.

Table of Contents

List of Tables

Introduction

Each year the Federal Aviation Administration (FAA) allots millions of dollars of its budget to hire and train Air Traffic Control Specialists (ATCS) (Gibb, 1991). Approximately 40 percent of these hirees are dropped from the FAA Academy due to the lack of skills that are deemed essential to the job of Air Traffic Control (ATC). Furthermore, attrition rates continue to be high after the ATCS candidates have been sent to the field. It is during this phase that an additional 11 percent of the new hirees are attrited (Della Rocco, Manning, & Wing, 1991).

The FAA currently uses a mass-testing instrument for candidate selection obtained through the Office of Personnel Management (OPM). This testing instrument is needed to maintain a selection system that is readily accessible to all segments of the population. However, this instrument is not adequate to ascertain a candidate's success in the field training phase. Prior to the 1976 introduction of the FAA Academy screen, the attrition rate in the field was 38 percent.

The FAA Academy screening process costs the government more than $10,000 per candidate (Broach, 1991). The Academy's central location and specialized staffing needs are major contributors to these high costs. By using a standardized computer-based test at various locations (for example, at the nine FAA Regional Headquarters Offices), the FAA may be able to reduce some costs. These cost savings may then be allocated to other areas within the entire FAA system.

1

Statement of the Problem

The purpose of this study is to evaluate the effectiveness of an ATC oriented computer-based screening test in predicting performance success when using an ATC computer-based simulator for training.

Review of Related Literature

The purpose of ATC is traditionally defined as providing the safe, orderly, and expeditious flow of air traffic. This entails three distinct, and sometimes incompatible, measures of system performance. ATCSs must adhere to rules and procedures in order to provide a fair and impartial service to all who may use the system (Hopkin, 1978).

In ATC, questions are posed, and answers expected in system terms. These may deal with the rate of traffic flow, handling capabilities, separation standards, communications (dealing with pilot/ATCS, intra- and interfacility channels), and staffing. These questions deal with the whole ATC system. System measures must relate to the abilities that the ATCS must possess to be successful within that system. The question is how does the FAA test, during the selection process, ATCS candidates to determine if a candidate will succeed as an ATCS.

Historical Development. The earliest efforts by the American Institute for Research (AIR) found several test measures to be predictive of ATC job performance. The major predictors were spatial ability, abstract reasoning, percepetual speed, mathematical reasoning, short-term and long-term memory, work sample measures, spatial-temporal reasoning, distance-temporal relationships, and encoding (Taylor, 1952).

The U. S. Air Force used the AIR findings to develop a battery of 37 tests (Brokaw, 1957). These tests found that personality indices were not valid predictors of performance. Brokaw's work was repeated by Cobb in 1960 (Cobb,1962). Cobb subsequently found strong support for tests centering on abstract reasoning, work samples, spatial ability, and mathematical reasoning. In 1964 these findings were used for both terminal and enroute trainees (Trite & Cobb, 1964). The results provided strong support for aptitude testing in ATC selection.

In 1961, the FAA began using testing for selection of ATCS candidates. Through research conducted by the FAA Civil Aeromedical Research Institute (now the Civil Aeromedical Institute,CAMI) the previous tests conducted by Brokaw led the Civil Service Commission (CSC) to establish a selection test for screening ATCS trainees. These test requirements for selection were implemented in 1962. This test battery remained basically unchanged until 1981. However, research concerning improvement of the battery continued.

Buckley, O'Connor, and Beebe (1969) attempted to relate system performance and individual performance measures. This study assessed whether system performance measures, which had been derived from a simulated ATC system for one person, could also be employed to measure the performance of the same individual ATCS's handling live traffic in a normal working environment. Performance with live traffic was assessed by anonymous ratings by peers and supervisors. This study achieved a higher reliability of performance ratings than is typical in ATC because the observers were thoroughly trained. However, the reliability of most ratings of controllers by other investigators has been low. This is due to the use of

subjective rating scales and much of the variance remaining unexplained. Henry, Kamrass, Orlansky, Rowan, String, & Reichenbach (1975) concluded that performance evaluation measures which do not rely on subjective ratings are essential in ATC.

It was during the latter half of the 1970s that a new initial selection test was developed. Research focused on studies that adhered to the Uniform Guidelines on Employee Selection Procedures. These procedures outline the development of field training performance ratings, the development of a longitudinal database for continuing validation research, and the refinement of the optimal combinations of old and new aptitude screening measures (Collins, et al., 1981).

A major study by Mies, Colmen, & Domenech (1977) attempted to predict the success of ATC applicants. It identified four areas of operational criteria for judging success. These areas were performance during training, performance on the job, progression according to the judged difference in complexity between originally assigned and current air traffic options, and continuation in employment as an ATCS. Additionally, a criterion of success was derived which included further factors which could be predicted by a test battery, by a pre-employment experience scale, or by an ATC occupational knowledge test, or any combination of these factors.

Research intended to improve the measures used to select controllers may add new measures to a test battery or to improve existing measures. Dailey and Pickrel (1977) adapted the simplified ATC task, as originally developed by Buckley, O'Connor, and Beebe (1969), for use in screening ATC applicants. This task measured the ability to detect conflicts. The Multiplex

Controller Aptitude Test (MCAT) Conflict and Aptitude scores produced

higher validities than any other tests then in use. The variables in the MCAT

Conflict and Aptitude test were prominent predictors in most regression

equations which predicted job performance success.

In 1986 Hunter and Schmidt, while working in Great Britain, developed a

computer-based aptitude test which was validated against pass/fail criteria for

initial training of Royal Air Force (RAF) ATCSs. After consideration of the

results, the computer-based test battery was instituted for the selection of

ATCSs in the RAF. This was the first use of computer-based testing

specifically for selecting ATCSs.

The U. S. Navy developed a computerized battery that included tasks to

evaluate static and dynamic spatial ability (Hunt, Pellegrino, Abate, Alderton,

Farr, Frick & McDonald, 1987). The static spatial tasks were tapping skills

that were measured as well by paper-and-pencil tests. The dynamic spatial

tasks appeared to require the additional ability of the use of dynamic

movement processing. This latter skill appears to be important for ATC, but

would be missed by any assessment procedures limited to obtaining applicant

information by paper-and-pencil.

Performance measurements. An approach to performance measurement

has been to treat the human as a system component and to describe the ATCS

in physical, mathematical or engineering terms that are compatible with other

system components (Hopkin, 1979). With the evolution of technology and the

introduction of new computer systems, the role of the ATCS has been

changing from an active controller to that of a passive monitor. The human as

a monitor can be described either in practical terms which are useful for

systems design or in terms of such psychological constructs as signal detection (Hopkin, 1979). This has implications for future selection procedures . The abilities that will be required of future ATCS's will change as the job description changes.

The range of measures of the ATCSs skills may be gauged by considering the ATCS handling heavy traffic in a familiar working environment. Measures of the system can verify that the traffic is heavy and that task load is high. Measures may also exist that establish that the system is safe and efficient (Hopkin, 1979). Measures of ATCS may include rates of activity, performance of tasks, and skill usage (which may include information processing and decision making, errors, delays, omissions and attention). Psychological or biochemical indictors may show the effort being made by the ATCS in handling the traffic safely.

Measurements of the individual ATCS are normally made as part of the initial selection procedure. These measurements are also used to assess the progress of the ATCS during training or retraining, or to establish the range of individual variability to be expected in task performance, or to gauge the impact of the system on the human (Evrard, 1975).

Current selection process. In the selection process, general intelligence seems to be the most important predictor of success. Some American psychologists appear to agree on three general domains of predictors. These contain: cognitive, which include mental and perceptual abilities; non-cognitive, encompassing personality traits and interests; and physiological or physical, embodying medical conditions and physiological indicators (Wing, 1991). Cognitive skills may be defined as those that require more internal

mental processing, such as reasoning and vocabulary. Non-cognitive predictors typically encompass personality variables such as the ability to handle stress, and a motivation to succeed. These items refer to feelings, attitudes, activities, and judgments and preferences about self and others.

The current ATCS selection process consists of two phases. The first phase of the selection process is the OPM ATC selection test battery. This is a paper-and-pencil format aptitude test battery consisting of three tests. These tests are classified as: the MCAT, Abstract Reasoning, and the Occupational Knowledge Test (OKT).

The MCAT was developed to establish a paper-and-pencil test with a higher predictive validity than tests included in the CSC selection battery. The design of the MCAT was intended to approximate simulated air traffic activities. Simulated airspaces are presented with several aircraft crossing the airspace. Candidates are required to identify situations that may result in conflicts between aircraft based on a limited set of separation criteria.

The Abstract Reasoning test measures two principles of logical development shown by sequences of numbers and letters (Collins, et al., 1981). The test has the candidate view two rectangular boxes. Within the first box are three figures with some common trait. Within the second box are two shapes with a common trait (different from the first box) and a question mark. The candidate must determine the relationship of the figures shown in the first box, then reason what the common trait is with the relationship of the figures in the second box. The candidate must then decide which of four possible answers fit the trait required by the question mark.

The OKT was developed to provide a more objective and reliable measure of a candidate's ATC job knowledge than was provided by the ratings that were provided by job history data (Della Rocco, Manning, & Wing, 1991). The OKT has no minimum cut-off score and it is utilized only to improve an applicant's ranking on the Federal Register. The OKT converts prior experience into a standardized measure to preclude the subjective evaluation of a candidate's background in the assigning of additional qualifying points.

In subsequent studies to assess the validity of OPM test scores in relation to field training performance measures, the MCAT battery scores were found to predict field training performance nearly as well as the FAA Academy grades (Manning, Della Rocco, & Bryant,1989; Della Rocco, 1990; Della Rocco, Manning, & Wing,1991). However, this conclusion was found to be valid only with respect to MCAT scores, and only predicting success for the enroute option candidates. The Abstract Reasoning and OKT scores did not predict any criteria for enroute field training. With respect to VFR Terminal candidates, the MCAT was found to be a poor predictor of performance. It was found that within the terminal option, only the OKT was a valid predictor of field training success. The MCAT provides for ATC situations that may appear within a rigid airspace structure along published routes. Conversely, the abstract reasoning and OKT tests tap a candidate's ability to project into a visual environment. As the OKT is based on predictive values of prior ATC experience, generally the only candidates with that experience were former military ATCS. However, it was found that most Academy academic and

laboratory results are better predictors of most measures of field training performance than the OKT.

Computer-based selection. Automation should have a major impact on the selection of the ATCS, and the management of the selection/ training system (Nyfield, 1991). Many nations still rely on paper-and-pencil testing devices for selection processes. The use of a computer-based selection device is becoming more attractive, especially as the job of the ATCS is becoming more automated. However, a note of caution should be issued. Computer-based testing systems are presently expensive, especially when compared against paper-and-pencil systems. The higher development costs of an automated system must show clearly defined benefits to justify the the use of a computer-based testing system. Automation allows for a greater choice of information and better use of the gathered information. It will be a benefit those who are involved in the selection process.

While pencil-and-paper testing is presently the more cost-effective, these tests can be adapted to computer-based testing. The computerized testing system would provide for greater standardization and ease of scoring. It will provide accurate results very quickly, thereby allowing for the immediate use of the information.

Selection is becoming a two-way process. The emphasis is on the employer discovering if a candidate will become a good controller and the candidate discovering whether the occupation is right for him/her. Automation can provide a potentially more cost effective way for a candidate to find about more about the role he/she would be exposed to in the ATC system. Using an interactive system that produces almost instant feedback to

the candidate allows the candidate to ascertain if he/she might be suited for the job. This could supply the employer with candidates who have "screened themselves" and have a better knowledge of what the job entails.

The literature suggests that automation will bring considerable change to ATC selection. Past research has demonstrated that performance-based instruments may be useful in predicting ATCS success in the training and operational environments. With the advent of lower cost computer systems in the marketplace, performance-based testing, using systematic methodology and validated procedures and criteria, will bring improved selection tools to aid in the selection of competent ATCS, and significantly reduce the costs of operating the National Airspace System (NAS).

Statement of the Hypothesis

Computers can process vast amounts of information quickly and accurately. A computer can measure cognitive abilities more effectively and accurately than a paper-and-pencil test. These abilities may be those that possibly are predictive of ATC success. A computer can measure more precisely such entities as reaction time of discrete events that are typical of ATC operations (paper-and-pencil tests at best only give an average response time per item). A computer can also provide a dynamic display of events/stimuli that is representative of ATC. This cannot be accomplished with paper-and-pencil tests.

It is for these reasons that computer-based tests may increase the predictibility of performance measures of an ATC simulation program used in

training. This assumption is correct if the abilities that need to be assessed cannot be measured by paper-and-pencil tests, and that these abilities are important in ATC training. It is therefore hypothesized that the utilization of the computer-based selection test battery will aid in the prediction of a candidates performance when using an ATC computer-based simulation program for training.

Method

Subjects

The sample was selected from the population of Embry-Riddle Aeronautical University (ERAU) undergraduate students. The subjects were enrolled in AT365, Air Traffic Control Operations & Procedures. Total enrollment for AT365 for the semester was 49 students. All students agreed to participate in the study. The incentive to participate was a variable number of points added to the course final grade (1-3 points after exam averaging). Of the total there were 38 students, 34 males and 4 females, who subsequently participated in the selection test battery. Appointed times were established for participation in the selection test battery. These appointments were made for times other than regular AT365 class time. Eleven students failed to appear at the appointed testing time or to reschedule an appointment. For identification purposes, each student that participated in the selection test battery used his/her social security number. It should be noted that because the FAA Academy is presently closed, it was not possible to test ATC candidates that were entering the actual FAA training program on the selection test battery.

AT365 utilizes a computer simulation program, TRACON/Pro™, developed by Wesson International, Inc.. This company is a computer software developer specializing in ATC training. This system is used at ERAU to teach the fundamentals of ATC in the ATC Minor. TRACON/Pro and its associated equipment simulate actual radar situations with sufficient fidelity. The presentation of the simulator is consistent with current ATC practices. The students enrolled in AT365 were introduced to ATC procedures and

12

operations for the first time. This lack of prior experience approximates the experience level of candidates that would be entering the FAA Academy ATCS training program. It is assumed, therefore, that the subject sample has an interest in the aviation field, but does not have any previous experience in ATC (prior military experience).

Equipment

The instrument that was used for this study was an experimental computer-based selection test battery that had been developed by researchers and designers at ERAU. The test battery was administered on a Gateway model 2000, 386-25 MegaHertz (MHz) personal computer (PC) system. The monitor used was a Gateway 2000, CrystalScan 1024 N1, 17" color monitor in association with a Gateway 2000 standard PC keyboard.

Test Battery. The test battery is comprised of four tests. These test are, in order of presentation, the mental rotation skills test, the absolute difference test, the verbal working memory test, and the spatial working memory test. Of the four tests only the first test and fourth test were utilized for this study. The rationale for using only the first and fourth tests were that these tests were those most closely associated with the ATC functions that were taught and measured in the ATC simulation exercises.

The first test evaluates mental rotation skills. This test is known as the manikin test and is an *ATC job related performance.* This is the skill necessary to mentally picture a situation, or screen, from different perspectives. This is accomplished by having the candidate mentally rotate a manikin. The manikin has shapes extending from both of the upper extremities. The figure is configured in various positions, facing the

candidate, turned away from the candidate (viewing the rear of the manikin), standing upright, or inverted (i. e., standing on its head). Any combination or variations may appear on the screen. The manikin will appear to be holding one of three different shapes of which two shapes are used, one in each of the upper extremities shown and a control shape illustrated at the top of the screen. The candidate is to identify which hand the control shape is positioned relative to the position of the manikin. There were eight trials with each trial having a one minute duration. All responses are evaluated for performance accuracy and reaction time. This test may be associated with the skill needed for the ATCS to mentally and visually rotate targets on a radar screen. This ability is required to properly direct an aircraft in a proper direction relative to its direction across the radar screen. This attribute was used with relation to proper direction commands (vectors) issued in the simulator exercise.

The second test evaluates the candidate's ability to store, manipulate, and recall short term memory functions at a rapid rate. The candidate is asked to respond by supplying the absolute difference between two numbers in succession. The third test is a verbal working memory task. Candidates are required to retain in their *working memory* the results of mental arithmetic while processing additional quantitative information.

The fourth test also deals with an *ATC job related performance* and is a test of spatial working memory. This test is also known as the grid test. The candidate is shown a grid that is 6 blocks wide and 6 blocks long. This is a two-dimensional presentation. Five aircraft representations are shown individually with associated direction and altitude information. These five

aircraft are identified as "A", "B", "C", "D", or "E". The identification letter for each aircraft appears in one of the blocks of the grid and the direction/altitude information is printed along the lower edge of the monitor screen under the grid presentation. An example of the text would be "Aircraft A at 3000 feet heading North". The aircraft's identification letter and corresponding flight information are presented for three seconds. This is done for all five aircraft letters. Upon completion, the numbers 1-6 appear at different positions on the grid. The candidate must determine which two aircraft have the potential to collide and at which location this will occur. Aircraft speed is not a factor in this test. The subject is instructed to declare a collision where the two flight paths cross at the same altitude. Accuracy and reaction time are the dependant factors for this test. This test consists of six trials of six problems each.

The minimum percent correct level for each of the tests is set at 70% for all of the tasks except for the final test which has no minimum. The purpose for these levels is to insure against random responses or the continuance of the tests without understanding of the instructions. If the subject does not achieve the minimum score, a failure notice for that test appears on the monitor screen and the test battery program terminates.

TRACON/Pro™. The final instrument necessary for this study is TRACON/Pro™. This equipment is a medium fidelity ATC simulation system. The system simulates the actual presentation of a terminal radar approach control (TRACON). Any TRACON facility may be represented by the program, whether simulated or actual.

The TRACON/Pro system was obtained for the purpose of instructing students enrolled in the ATC minor. This equipment utilizes standard 486, 33 MHz personal computers. The equipment is installed in custom-made cabinets that simulate ATC radar workstations. Each workstation consists of a 21" color monitor (the radar presentation), a 14" monochrome monitor that displays weather reports and associated information, a flight progress strip bay, a back-lighted overhead display with airspace representations, a standard PC computer keyboard, and an ARTSIIIA keyboard. This latter keyboard is a representation of an actual radar input device used by ATCSs at various TRACON locations around the country.

The program has an integrated scoring system which evaluates the candidate's performance with a numerical score. This score is calculated using the total amount of possible points for the scenario. This point total is based on the complexity of the scenario, the aircraft flight environment (arriving, departing, or enroute), and weather parameters (instrument meteorological conditions or visual meteorological conditions). The scoring values are provided in Appendix C.

Any point errors made by the candidate are deducted from the total possible points. These errors are deducted using values of severity based on the FAA Handbook 7110.65. This book contains the standards and rules that the ATCS must follow to properly separate air traffic. Point values are assigned to actions that are contained within two categories designed into the simulator scoring system. One category involves the command inputs. This category records all of the commands that are issued by the student during the scenario. A list of commands is shown in Appendix D.

The second category designed within the simulator scoring system is the recording of errors made by the student during the scenario. These are errors that are the consequence of commands issued to the aircraft. A list of errors is shown in Appendix E.

Design

The correlation method of research was utilized to conduct this study. Performance scores from both the selection test battery and the ATC simulation program, TRACON/Pro, were correlated in order to determine the relationship of the test battery scores to the performance score on Tracon/Pro.

Task stability, performance learning curves, and the magnitude of individual differences have been established on the test battery and are suitable for this type of research. Forty subjects were administered the test battery. The range of ATC experience of the subjects included persons with extensive ATC experience, moderate to little ATC experience, and/or other aviation related experience and included individuals with no prior involvement in the aviation field.

The relationship evaluated would be between the different parts of the selection test battery and similar parts of the simulator scoring. The mental rotation skills test (manikin test) would be equated to the vectoring commands structure and the spatial working memory test (grid test) would be equated to the separation error structure of the simulator program.

Procedures

This study continued to evaluate the effectiveness of the experimental computer-based selection test battery to predict a subject's performance in controlling air traffic when using an ATC simulator. The results of performance scores from the selection test battery and the performance scores from the ATC simulator program (TRACON/Pro) were obtained from each subject.

Each student was given the test battery prior to any instruction about the ATC simulator. A controlled environment was provided in order to ensure that all subjects would have an equal chance to perform to their highest possible level on the test battery. The subjects were tested in a quiet room. Access to the room was limited to the examiner and subjects in order to minimize any distractions. Each subject was briefed about the test battery. Several demographic data questions were entered into the database to establish identity. The subjects were then told to read the instructions for each task and then to proceed through each test. Instructions were all presented by the computer and the pace was controlled by the subject.

The subjects received approximately 12-15 hours of classroom lecture during normal class meeting times. The classroom lectures explained basic practical applications of the rules and regulations contained in the FAA Handbook 7110.65. The regulations reviewed were those regulations concerned with lateral, vertical, and horizontal separation standards. These regulations formed the foundation for keeping aircraft properly separated. Subsequent class lectures taught basic phraseology including the phonetic alphabet, number usage, number grouping and aircraft callsigns. Other

lectures taught the students about the airspace that was to be used for the simulations. These lectures explained the runway locations, headings, approach procedures, and the various instrument landing procedures for a particular runway. All topic areas were tested to ascertain that the students were receiving adequate instruction.

Each student then received nine hours of practice on the simulator prior to taking the final performance test. All instruction was issued during the regular class hours for AT365. All candidates performed and practiced on the same scenarios and received the same information from the researcher and assistant. Nine hours of practice had previously been determined to be sufficient. This time length was determined by the researcher to be sufficient through practical experience as an air traffic controller in the training of ATCS developmentals. All subjects performed the same scenarios during the instructional phase and performed the same final evaluation scenario. The researcher and assistant administered all instruction on the Tracon/Pro™ simulator.

The first two hours of simulator instruction was used to introduce the students to the simulator. This instruction centered on the fundamentals of the system and the presentation. This included the use of the different keyboards and trackball, correct keystroke entries to activate a scenario and to control the targets. This was necessary due to the fact that two distinct keyboards are used by the simulator. One keyboard is a standard PC keyboard, and is used when usiing the stand-alone mode of training. The standard PC keyboard is also used by a pseudo-pilot to input commands received from the student controller. The second keyboard is a functional

replica of a radar workstation keyboard used by the ATCS to input different information into the ATC computer system. This second keyboard is different from a standard PC keyboard and instructional time must be alloted to make students familiar with its functions. This time was also used to instruct scope orientation (North), runway and navigation fix recognition and presentation.

The next two hours were used to instruct students in vectoring of aircraft. These sessions utilized a simple maze (designed by the researcher) with an airway centerline to help teach procedures that keep aircraft on airways and within boundaries. The students then practiced vectoring aircraft within the airspace boundaries.

The remaining four hours was used to instruct students, and have them practice vectoring aircraft to an instrument approach, departure procedures, and separation of aircraft within the airspace jurisdiction.

TRACON/Pro has a built-in scoring system which evaluates the student's performance and gives a numerical score. The score is calculated by starting the scenario with a preset point value based on the complexity of the scenario. This is determined by the computer. Points are awarded to the student for tasks that are successfully accomplished. Points are subtracted for each separation/ control error that is made. The total number of points received for completing all of the tasks is summed. This is then computed against the penalty points and the result is presented to the student. This system was used to assess the performance of the students.

The amount of air traffic and flow was consistant with the amount of instruction given to the students in the time allotted. The scenario was

designed as a modification of the parameters used by the FAA for its

performance verification for a Level III approach control facility.

The selection test battery performance scores were correlated with the

TRACON/Pro scores. The subsequent regression coefficient indicts the degree

of relationship between the scores.

Analysis

The purpose of this study was to continue to evaluate the predictive value of the Embry-Riddle designed computer-based ATCS selection test battery. The abilities assessed by the selection test battery were believed to successfully predict ATC performance on a simulator, based on findings from previous studies.

The computer-based ATCS selection test battery consists of four tasks. These tasks include measures for mental rotation abilities, short-term memory capabilities, verbal working memory capacities, and spatial working memory capabilities. Each task has two dependent variables; that of correct reaction time (CRT) and amount (percentage) correct. In a prior study, and during the design of the test battery (Gibb, et al., 1991), the data indicated the number of trials that were necessary for a subject to reach a stable level of performance for each task. Each task has multiple trials. The first and second tasks (mental rotation and short-term memory capabilities) have eight trials each. Of these only the final four trials are used for data. It is after the fourth trial that the learning performance curves were shown to reach a stable level. The learning performance curves for the last four trials of each of these two tasks was shown to indicate only a small linear improvement in performance. The third tasks contains three trials. Only the final two trials are utilized to evaluate the verbal working memory capacity. The fourth task contains six trials. All six are utilized for analysis in this study.

Two factors were established for each of the two tasks. One is for percent correct answers, the other for the correct reaction time. The means were calculated by collapsing the data for only the stable trials of each task.

From previous studies, it was expected that a moderate correlation (R=+.30 or greater) would be found between the test battery and the performance results from the simulator program. A limited relationship was tested between two tests of the selection test battery and two scoring areas of the simulator program. The first was between the manikin test and the vectoring commands of the simulator. The second was between the grid test and the conflict error scores. The following tables illustrate the results:

Table 1

Mean and Standard Deviation of Performance Scores and Test Battery Scores

	Mean	Standard Deviation
Performance Scores-Vectoring (TRACON/Pro)	14.03	7.11
Manikin Test	90.18	26.51
Performance Scores-Conflict Errors (TRACON/Pro)	0.92	1.22
Grid Test	62.63	18.49

Table 2

Regression Analysis Versus Tracon/Pro™ Performance Criteria

Performance Task	Multiple R	Multiple R^2
Manikin Task Accuracy	.037	.001
Grid Task Accuracy	.183	.033

As noted from the above tables, only accuracy measurements were utilized for this study. The CRT function of the test battery program was adulterated by the corruption of certain encoding sequences. It is believed that the source code for the test battery became partially disabled during the transfer from the master system to the user system. CRT measures are integral to the test battery program. The CRT task measures can not be separated from the battery as their counterpart measure does contribute to predicting performance. Therefore, without the corresponding CRT connections, the accuracy measurements are one-sided and cannot provide an accurate picture of the students selection potential.

Conclusions

This study was conducted to continue to assess the effectiveness of a computer-based selection test battery. The selection test battery was developed to aid in predicting the performance of an ATC candidate. The purpose for the selection test battery is not to predict a subject's performance on the TRACON/Pro simulator. In order to augment that effectiveness with computer-based simulator training, the simulator that is used to conduct the performance measurement must contain a software program that is consistent with the needs of the evaluation process.

The first part of the experiment was the administration of the computer-based selection test battery. There were deficiencies in this software, the major deficiency being the loss of the reaction time measures. As these measures are crucial to the complete scoring process as reaction time is needed to respond quickly and properly in any given ATC situation. The lack of these measures only produce a partial results and therefore cannot accurately indicate any predictive tendencies. The loss of the reaction time measures was not discovered until the completion of the selection test battery by the students. When this fault was found, it was too late to redesign the experiment due to the time constraints.

Only a limited number of abilities were tested. In the field of ATC there are more than four abilities needed by an ATCS in order to properly function within the ATC system. While the test battery can predict certain quantitative abilities, there are other qualitative abilities that are needed by the ATCS.

Student motivation may also have been a factor affecting the results of this study. Though an incentive was offered by the investigator in the form of additional bonus points to their grades, student participation in all phases of the experiment was 77.5% (38 of a total of 49). This may have been the result of the belief that the results of this test battery did not really count for anything (for example, the knowledge that the FAA was not hiring ATCS candidates).

Another reason for the lack of relationship of test battery scores to performance scores appears to be the weak scoring structure of the TRACON/Pro software program. The scoring system incorporated on TRACON/Pro is an evolution of the computer game, Tracon for Windows™ produced by Wesson International, Inc. While useful for gaming purposes, it does not appear to be dynamic nor robust enough to be used for actual performance verification. Due to the above mentioned deficiencies, the stated hypothesis of this study could not be tested.

A more accurate assessment of test battery scores to predict ATC performance scores may be accomplished using standard FAA over-the-shoulder evaluations. However, the use of the over-the-shoulder evaluation method may result in highly subjective evaluations. It may cause a serious bias using this evaluation method, the over-the-shoulder evaluations in this instance would have been conducted by the researcher who administered the test battery.

It has been observed by the researcher that, in lieu of the built-in computer managed training (CMT), test battery scores may be correlated, to a degree, with student performance using over-the-shoulder evaluation

methods. The present ATC simulator laboratory in use at Embry-Riddle Aeronautical University is a medium fidelity representation of a FAA TRACON. The instructors, including the researcher, of the various ATC classes are considered as ATC subject matter experts (SME). Both laboratory class instructors are full performance level (FPL) controllers from different facility environments in the FAA and Europe. However, using the over-the-shoulder method of evaluation, is labor intensive.

However, on-the-job evaluations in the academic environment are impractical. As mentioned previously this type of evaluation method is labor intensive. It requires a 1:1 ratio of student to instructor. Based on this requirement, a class with 30 students would need 30 instructors. Presently there are two instructors for 49 or more students. This results in at least a 25:1 ratio. This ratio shows that one instructor cannot properly attend to the needs of any one student except for a short time. In order for this evaluation to be effective, the student must be monitored and evaluated over a long duration. This time length of time is not available to the instructor, and therefore the simulator scoring system is essential to the evaluation of student performance. The weakness of performance scoring of the TRACON/Pro system was not known prior to this study. As presented, the system scoring system was fully functional. The resulting low correlation coefficient may in part be due to the scoring profile of the TRACON/Pro scoring system not accurately reflecting performance. This was illustrated directly from the raw data.

Improving the quality of training that a student controller receives before entering the real-life work environment would have a significant impact on

the above mentioned problem. Studies are still in progress as to the validity of simulator training in ATC. As the use of computers are becoming more commonplace in the field of training, the inclusion and reliance of internal performance scoring becomes more important. Strong performance parameters are essential to properly assess a students ability. In this area the TRACON/Pro system appears to be deficient. It is this deficiency that appears to to result in a low correlation between the selection test battery and performance scores. The performance scoring system is based on an arbitrary set of values that may or may not be based on any useful data. It sets quantitative values on artificial parameters that are not realistically portrayed. The performance scores produced by the TRACON/Pro system do not accurately represent a student's performance. There has not been an evaluation of the relationship between the over-the-shoulder evaluation parameters and the TRACON/Pro performance system.

Another important factor of the value of the simulator scoring system was its' lack of reliability. This was discovered only after the experiment was completed. For example, some command results were accurate as to the number of commands given for some students while other results showed no commands were given at all. These results occured in a random manner.

The predictive value of the selection test battery appears to show a greater correlation to performance scores with the use of over-the-shoulder evaluation methods. However, this may be due to the researcher's and the SME's familiarity with the simulator and experience in using over-the-shoulder evaluations.

Previous studies have indicated that there is a correlation between the selection test battery and performance predictions. Further research in this area should be conducted in order to formulate an effective CMT if the trend to computer simulators continues to expand.

Recommendations

Any computer-based selection test battery programs must be relevant to their given task. Additionally, the need for effective CMT is essential for a simulator system to be effective in training future ATCS. The following recommendations are made based on the results of this study.

There should be a greater range of tests in the test battery to better predict the success of a potential ATCS candidate. The present test battery only touches on some of the abilities believed to effectively predict success. The test contains two sections that are correlated most directly with ATC functions, the manikin test and the grid test. However, these tests are two dimensional. They are portrayed on a flat screen that is representative of a radar screen. It should be noted that the FAA has more Air Traffic Control Tower (ATCT) facilities than there are radar facilities. The tower environment (both VFR ATCT and IFR ATCT) relies on visual separation of traffic both airborne (separate and sequence aircraft in the local traffic pattern) and with aircraft arriving and departing the airport surface. This visual separation ability is a predominent expertise that is needed by the tower ATCS. This ability should be incorporated into the test battery as a function of depth perception with accuracy and reaction time measures.

Another skill that should be tested is the ability to prioritize. This skill is required by every ATCS regardless of the type of facility. While the present test battery may indicate how a task is done, there is no indication of what importance is placed on the sequence that the task is to accomplish. A

priority test could indicate how a candidate would solve a series of events and effect a successful conclusion.

The results of this study also suggest a definite need to develop a good and realistic computer-generated performance scoring system. The present system on the TRACON/Pro simulator is based on arbitrary, and for the most part unrealistic, scoring parameters. This does not reflect the full spectrum of ATC aptitudes, priorities and other any variables of ATC. Computer simulators measure objective results based on written rules and regulations that form the basic structure of a given occupation. This is true in the field of ATC. The basic rules and regulations used by the ATCS are contained in the FAA Handbook 7110.65. However, within those rules is the freedom, not to break the rules but to "adapt" those relevant rules to fit the situation. These are the subjective areas that need to be developed and incorporated in a computer generated performance scoring system. By combining the objective parameters and the subjective abilities, a true performance scoring system can be developed.

After the performance scoring has been developed it will be necessary to validate the computer-generated performance evaluations against other performance measures. This may be partially accomplished using the present over-the-shoulder evaluation form as used by the FAA. This form contains both objective evaluation measures (letters of agreement, separation standards, equipment status, and capabilities) and subjective evaluation measures (control actions, control judgment, awareness). This form could provide a basic framework from which to develop and validate a good, operational computer-generated simulation performance scoring system.

References

Boone, J. O. & Lewis, M. A. (1979). *The selection of air traffic control specialists: Two studies demonstrating methods to insure an accurate validity coefficient for selection devices* (Report FAA-AM-79-14). Oklahoma City, OK: FAA Civil Aeromedical Institute.

Broach, D. (1991). *Personal communication with Gibb.*

Buckley, E. P., O'Connor, W. F., & Beebe, T. A. (1969). *A comparative analysis of individual and system performance indices for the air traffic control system* (Report FAA-NA-69-40). Atlantic City, NJ: FAA National Aviation Facilities Experimental Center.

Cobb, B. (1962). Problems in air traffic management: II. Prediction of success in air traffic controller school. *Aerospace Medicine*, 33, 702-713.

Collins, W. E., Boone, J. O., & VanDeventer, A. D. (Eds.). (1981). The selection of air traffic control specialists: History and review of contributions by the Civil Aeromedical Institute, 1960-1980. *Aviation, Space, and Environmental Medicine*, 52, 217-240.

Dailey, J. T., & Pickrel, E. W. (1977). *Development of new selection tests for air traffic controllers* (Report FAA-AM-77-25). Washington, DC: FAA Office of Aviation Medicine.

Della Rocco, P. (1990). *Selection of air traffic controllers for automated systems: Applications from current research.* FAA-TR-13. Washington,DC: Federal Aviation Administration.

Della Rocco, P., Manning, C. A., & Wing, H. (1991). Selection of air traffic controllers for automated systems: Applications from today's research. In J. A. Wise, V. D. Hopkin, M. L. Smith (Eds.) *Automation and Systems Issues in Air Traffic Control.* Springer-Verlag, Berlin, Germany.

Evrard, E. (1975). Medical problems relating to air traffic control personnel. In A. Benoit (Ed.) *A survey of modern air traffic control.* Neuilly-sur Seine, France: ADAR Dograph No. 209.

Gibb, G. D., Smith, M. L., Swindells, N., Tyson, D., Gieraltowski, M. J., Petschauer, K. W., Haney, D. U. (1991). *The development of cognitive tasks for the selection of air traffic control personnel* (Technical Report No. CAAR-15426-91-7). Daytona Beach: Embry-Riddle Aeronautical University, Florida, Center for Aviation/Aerospace Research.

Henry, J. H., Kamrass, M. E., Orlansky, J., Rowan, T. C., String, J., & Reichenbach, R. E. (1975). *Training of U. S. air traffic controllers* (Report FAA-AD/A-006603). Washington, DC: Office of Personnel and Training.

Hopkin, V. D. (1978). An appraisal of real-time simulation in air traffic control. *Journal of Educational Technology Systems*, 7(1), 91-102.

Hopkin, V. D. (1979). Mental workload measurement in air traffic control. In N. Moray (Ed.), *Mental Workload*. New York: Plenum.

Hopkin, V. D. (1980). The measurement of the air traffic controller. *Human Factors*, 22, 547-560.

Hunt, E., Pellegrino, J. W., Abate, R., Alderton, D. L.,Farr, S., Frick, R. W., & McDonald, T. P. (1987). *Computer-controlled testing of visual-spatial ability*. (Report NPRDC-TR 87-31). San Diego, CA: U. S. Navy. Personnel Research and Development Center.

Hunter, D. & Schmidt, V. (1986). A computer based testbattery for selecting fighter and air traffic controllers. *Memo No. 8/86*. London, England: Director Science (AIR), Ministry of Defense.

Manning, C., Della Rocco, P., & Bryant, K. (1989). *Prediction of success in FAA air traffic control field training as a function of selection and screening test performance*. FAA-AM-89-6. Washington, DC: Education and Public Affairs.

Mies, J. M., Colman, J. G., & Domenech, O. (1977). *Predicting success of applicants for positions as air traffic control specialists in the air traffic service* (Report FAA Contract DOT FA75WA-3646). Washington, DC: Education and Public Affairs.

Nyfield, G. (1991). Automation issues for the selection of controllers. In J. A. Wise, V. D. Hopkin, M. L. Smith (Eds.) *Automation and Systems Issues in Air Traffic Control*. Springer-Verlag, Berlin, Germany.

Saville & Holdsworth Ltd. (1988). *Civil Aviation Authority predictive validation study. The selection of air control officer cadets. Supplemental report: RAF selection data*. Esher, England.

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

Trites, D., & Cobb, B. (1964). Problems in air traffic management: IV. Comparison of pre-employment, job related experience with aptitude tests as predictors of training and job performance of air traffic control specialists. *Aerospace Medicine*, 36, 1131-1138.

Wesson International. (1993). Instructional and Support Manual. *Tracon/Pro™: Professional air traffic control training systems*. Austin, TX.

Wing, H. (1991). Selecting for air traffic control: The state of the art. In J. A. Wise, V. D. Hopkin, & M. L. Smith (Eds.) *Automation and System Isssues in Air Traffic Control.* Springer-Verlag, Berlin, Germany.

APPENDIX A

AT365 AIR TRAFFIC CONTROL OPERATIONS AND PROCEDURES

AT365 is a basic course in the procedures and techniques used by air traffic controllers to ensure the safe, orderly, and expeditious flow of air traffic. This course consists of both traditional classroom (lecture/discussion) work and performance based instruction using an air traffic control radar similator. The airspace used in the simulator is represented by the Daytona Beach Terminal Radar Approach Control class C airspace.

The course is designed to provide the student with an opportunity to perform air traffic control tasks and work within a simulated environment. The student has an opportunity to experience air traffic control on a personel, "hands-on" basis, and becoming familiar and gaining an appreciation for the demands of the air traffic control profession. The student will gain an appreciation of the role of air traffic control as an integral part of the National Airspace System and an insight into the support systems and structure involved in the operation of an air traffic control facility.

APPENDIX B

FAA LEVEL III TERMINAL APPROACH CONTROL PERFORMANCE

VERIFICATION

The Performance Verification (PV) scenario of events as defined by the FAA Level III TRACON are as follows:

Arrivals; Primary Airport,

Minimum 10 aircraft, Maximum 18 aircraft

Mix of aircraft types and operating characteristics

Include aircraft departing from a satellite airport landing at the primary airport

Departures; Primary Airport,

Maximum 10 aircraft

Arrivals; Satellite Airport

Two arrivals adequate

Departures; Satellite Airport

Two departures adequate timed to fit in with departure flow

Overflights;

Two General Aviation types at mid-level altitudes

VFR aircraft;

Mix of primary targets, transponder-equipped, Mode C

Weather;

Basic VFR conditions

APPENDIX C

TRACON/PRO™ SCORING

The scoring variables installed in the Tracon/Pro™ simulation system:

Add:

enroute aircraft	500
departure aircraft	500
arrival aircraft	800
aircraft type	100

Subtract:

missed approach	250
aircraft off radar	1500
less than 3 mile separation	1000
less than 1 mile separation	5000
aircraft crash or mid-air collision	10000

APPENDIX D

TRACON/Pro™ TOTAL COMMANDS LIST

The Total Command List structure presented on the simulator system:

Total Commands	(Number)
Vectoring Commands	(Number)
Altitude Commands	(Number)
Speed	(Number)
Instrument Appreoaches	(Number)
Visual Approaches	(Number)
Say...	(Number)
Information	(Number)

APPENDIX E

TRACON/Pro™ TOTAL ERRORS LIST

The Total Errors List structure presented on the simulator:

Total Errors	(Number)
Crashes	(Number)
Separations	(Number)
Vectoring Errors	(Number)
Altitude Errors	(Number)
Handoffs	(Number)
Missed Approaches	(Number)