The Development of an Active Listening Test for Federal Aviation Administration Air Traffic Control Specialists

M. Kathryn Bleckley* and Cheryl L. Wilkinson*
*Civil Aerospace Medical Institute, Federal Aviation Administration, Oklahoma City, Oklahoma, USA

1. Introduction

The Federal Aviation Administration (FAA) is responsible for the safe and expeditious movement of air traffic in the United States. To meet this standard, the FAA hires and trains Air Traffic Control Specialists (ATCSs). These ATCSs maintain separation between aircraft from “gate to gate.” Air Traffic Control Tower (ATCT) controllers ensure aircraft are separated on the taxiways, runways, and in the air immediately surrounding an airport. Terminal Radar Approach Control (TRACON) specialists sequence aircraft into and out of the ATCT airspace, and Air Route Traffic Control Center (ARTCC, also known as en route) controllers also use radar to maintain separation during the high and fast portions of a flight.

An important part of a Human-Systems Integration effort is not only the human factors of design, but how the training and selection for that design are structured (e.g., Durso, Boehm-Davis, & Lee, 2015). The FAA training program for ATCSs begins at the FAA Academy with courses in air traffic basics and initial qualifications for each type of control facility. Graduates continue their training at their assigned field facility. Total training time varies by type and complexity of facility, and ranges from nearly 18 months to more than 3 years. Thus training represents substantial costs to the FAA.

These costs are one factor motivating the FAA to develop and use valid and fair selection methods. The FAA has used several methods (see Manning, Kegg, & Collins, 1988; or Heil & Reese, 2002, for reviews of this history). The current method includes the Air Traffic Selection and Training (AT-SAT) test battery. AT-SAT was developed to assess abilities and skills identified in the Nickels, Bobko, Blair, Sands, and Tartak (1995) job task analysis. Seventeen candidate tests were developed and evaluated for use in the final version of the AT-SAT battery. Of these, eight were selected for use: Dials, Applied Math, Scan, Angles, Letter Factory, Air Traffic Scenarios Test, Analogies, and the Experience Questionnaire (Ramos, Manning, and Heil, 2001a). The Experience Questionnaire was dropped from the battery in 2014.

There are three dynamic tests: Scan, Letter Factory, and Air Traffic Scenarios Test. Scan presents from 2 to 12 targets that move across the screen in 1 of 8 directions. Each target has an identifier and a value. If the value is outside a stated range, the applicant enters the identifier via the numeric keypad. The stated range changes without warning during the test. In Letter Factory the applicant "works" in a just-in-time factory loading letters (A,B,C,D) into boxes according to specified rules. When the scenario ends, the applicant then answers situation awareness questions. The Air Traffic Scenarios Test is a low-fidelity simulation of a controller’s job that tests the examinee’s ability to learn and follow simple procedures, visualize and project paths in 3 dimensions, monitor several objects at once, plan ahead, and remember to execute elements of the plan at the appropriate time.

The remaining 4 tests (Dials, Applied Math, Angles, and Analogies) are not dynamic, and therefore, could be administered with paper and pencil; however, the use of computers to present the items allows for timing the tests and collecting additional performance measures. The Dials test measures how quickly and accurately examinees can read dials on an instrument panel. The Applied Math test measures the examinee’s ability to apply mathematics to solve problems involving the traveling speed, time, and distance of aircraft. The Angles test measures the ability to recognize angles. Some questions present a picture and the applicant must choose the correct number of degrees from 4 choices; others present number of degrees and the applicant must choose the correct pictorial representation. The Analogy test assesses inductive reasoning and information processing skills using 3 analogy types: Non-Semantic Words, Semantic Words, and Non-Semantic Visuals.

These tests assess 8 of the top 10 worker requirements identified by Nickels et al. (1995) including recall from interruption, prioritization, tolerance for high intensity, compose, situation awareness, planning, execution, and thinking ahead. Oral communication, the ninth worker requirement, is assessed during an interview. The only worker requirement in the top 10 not currently assessed is active listening. Active listening, as defined by Nickels et al. (1995), is “the ability to hear and comprehend spoken information. This
ability requires an individual to recognize/pick out pertinent auditory information” (p.68). This must be accomplished in a noisy environment where many voices are speaking at once. Additionally, it is frequently necessary that active listening be accomplished while performing other tasks, such as scanning the control environment or entering data.

Components of active listening, such as selective attention, have been studied in controlled laboratory environments, thus providing researchers with a solid foundation. However, there is little work on active listening as an intact construct relevant to an ecologically valid context (e.g., ATC).

We have developed a potential assessment of active listening that may prove useful in predicting training success in ATCSs. This assessment, called the Auditory Attention Test (AAT), is patterned after the “Ready Charlie” task developed by Bolia and colleagues (see Bolia, Nelson, Ericson, & Simpson, 2001), in which faux pilots listen for a cue (“Ready”), their call sign (“Charlie”) which is then followed by brief instructions (“Go Blue 3”). The pilot then touches a colored square and a digit on a touch screen monitor. The critical information is embedded in a sequence of distractor instructions to other pilots. However, we wished to develop a task that more closely resembled the ATCS work environment in which several competing strings of auditory information may be present simultaneously.

2. Methods

Stimuli were developed from the Harvard Sentences (Egan, 1948; IEEE, 1969). First, comprehension questions with appropriate distractors were developed for most sentences. For example, if a sentence was “The girl was wearing a red dress.” (not a real stimulus), the comprehension question might be “What color was the girl's dress?”, then a distractor might have been “blue.”

Recordings of 7 females and 7 males reading the sentences were made. We developed a computer program to deliver the auditory stimuli at up to 7 virtual locations placed around the participant, subtending an arc of 180°, as shown in Figure 1. Each auditory location read a different sentence articulated by a different human speaker. Gender of the speakers, location of speakers, and sentences were randomly chosen for each trial. All participants received the same stimuli in the same order.

The program also presents a visual tracking task as a secondary task to provide greater realism. The computer screen in a typical trial is represented in Figure 2. The participant’s task is to attend to the sentences presented at the cued location while simultaneously keeping the line within the rectangle by pressing the ‘z’ and ‘/’ keys as needed. Following each set of sentences, the software delivers and scores the comprehension questions.
3. Results

Data collected from FAA Air Traffic Academy students (N=1,388) were analyzed. Thirty-seven participants gave no correct answers to the comprehension questions and performed perfectly on the visual task. Their data were excluded from further analyses.

Overall, as seen in Figure 3, performance on the comprehension questions was poor (N=1351; mean proportion correct = .36; minimum = .06, maximum= .70), suggesting that the initial configuration of the test was too difficult. Figure 3 shows that none of the participants scored more than 70% correct. Further, only 7% were able to get at least half of the questions correct. In fact, 14% scored at chance (i.e., .25) or below. The fact that the distribution was reasonably Gaussian suggests that the test itself should be fair once the overall mean is corrected. Ideally, a mean of .7 would bring this test in line with other AT-SAT tests assuming we retain similar variance.

Figure 3. Distribution of mean proportion correct for the comprehension questions. The red bar represents chance performance.
Additionally, strong performance on the secondary tracking task, as measured by mean seconds outside the box, suggests that the secondary task was taking precedence for many participants. Mean time outside of the box was .73 seconds. Average time for a trial was 12 seconds; thus, the average participant was able to track accurately 94% of the time.

We will present these results and will also present results of analyses of subsequent versions of the test. We will discuss at each opportunity how changes helped bracket on a version that can successfully be added to the selection and training battery.

4. DISCUSSION

Because the auditory task comprehension was poor and the visual tracking task performance was good, we adjusted several experimental parameters. These included eliminating 2 voices that were not perceptually similar to the others, fixing the locations from which the 4 voices are heard (0°, 60°, 120°, and 180°) while still assigning the location to be attended through randomization, and eliminating sentences with archaic language. We will present the results from this new version and any subsequent versions at the conference.

Tracking and reporting the development of selection instruments is an important but often forgotten part of the process of selection test development. By reporting test failures, as well as successes, we should be able to converge more quickly on an appropriate battery of tests.

Acknowledgments

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

References


