DOT/FAA/AR-99/42

Office of Aviation Research
Washington, DC 20591

Test and Evaluation Plan for Determining Screener Training Effectiveness

PB99-149692

Brenda A. Klock
J. L. Fobes, Ph.D.

Aviation Security Human Factors Program,
AAR-510
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

April 1999

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No. DOT/FAA/AR-99/42	2. Government Accession No.	3. Recipient's Catalog No.
4. Test and Evaluation Plan for Determining Screener Training Effectiveness		5. Report Date April 1999
		6. Performing Organization Code AAR-510
7. Author(s) Brenda A. Klock J. L. Fobes, Ph.D.		8. Performing Organization Report No.
9. Performing Organization Name and Address U. S. Department of Transportation, Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405		10. Work Unit No. 11. Contract or Grant No.
12. Sponsoring Agency Name and Address U. S. Department of Transportation, Federal Aviation Administration Associate Administrator for Civil Aviation Security, ACS-1 800 Independence Avenue, S. W. Washington, D. C. 20590		13. Type of Report and Period Covered 14. Sponsoring Agency Code ACS-1

15. Supplementary Notes: Draft Report Prepared By:

Joshua Rubinstein, Ph.D. & William Maguire, Ph.D.
Federal Data Corporation (FDC)
Science and Engineering Division
500 Scarborough Drive
Egg Harbor Township, NJ 08234

16. Abstract

The efficacy of Computer-Based Training (CBT) programs potentially useful for security checkpoint screener training will be evaluated at three different airports. Candidates will be trained with one of four CBT programs and the Screener Readiness Test, designed to assess screening-related knowledge, will then be used to evaluate the effectiveness of the CBT programs.

17. Key Words Computer Based Training (CBT), Screener Readiness Test (SRT), Aviation Security	18. Distribution Statement This document is available to the public through the National Technical Information Service (NTIS), Springfield, Virginia 22161		
19. Security Classif. (of this report) Unclassiifed	20. Security Classif. (of this page) Unclassified	21. No. of Pages 10	22. Price

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

CONTENTS

ILLUSTRATION

Table **Page**

ACRONYMS

ACSSP	Air Carrier Standard Security Program
ATL	Atlanta-Hartsfield International Airport
CBT	Computer-Based Training
COIC	Critical Operational Issues and Criteria
DTW	Detroit Metropolitan Airport
FAA	Federal Aviation Administration
HFE	Human Factors Engineers
MOE	Measure of Effectiveness
MOP	Measure of Performance
OJT	On-the-Job Training
OSM	Overall Success Measure
OT&E	Operational Test and Evaluation
SEA	Seattle-Tacoma International Airport
SRT	Screener Readiness Test

1. INTRODUCTION

1.1 Background

The effectiveness of the national civil aviation security system is highly dependent upon the people who are employed as checkpoint screeners. The training of these individuals is critical to their performance on the job. The Federal Aviation Administration (FAA) is very interested in enhancing screener training and further improving their readiness for the job.

According to Federal Aviation Regulations § 108.17 (Use of X-ray systems), there shall be a program for initial and recurrent training of operators of X-ray systems that includes training in the efficient use of X-ray systems and the identification of weapons and other dangerous articles. Section XIII of the Air Carrier Standard Security Program (ACSSP) presents the standards for training and testing of persons performing screening and security functions. For many years, the only FAA-approved training was that developed by the Air Transport Association. This 12-hour initial screener training program includes 40 multiple choice questions and 40 X-ray images to assess mastery prior to On-the-Job Training (OJT). In April 1997, the FAA also approved the use of a Computer-Based Training (CBT) system for initial screener training prior to OJT. This training is also based on Section XIII of the ACSSP. There are other training systems currently being developed for initial screener training. These systems have not been tested in the United States to see if they meet the standards for initial screener training.

As additional training systems are offered for initial screener training, each is expected to include a different test to assess mastery prior to OJT and screener certification. Because the variety of training options is growing, the FAA is developing a single uniform measure of mastery of initial training, the Screener Readiness Test (SRT) (Fobes, Neiderman, & Klock, 1999). The SRT contains X-ray images to be resolved for threat articles such as improvised explosive devices, the FAA's modular bomb set, hand grenades, guns, and knives. This preparedness evaluation also contains multiple choice questions on the major checkpoint screening tasks of walk through and hand-held magnetometer, pat downs, hand searches, X-ray operation, trace detector operation, and monitoring the exit lane. Once validated, the SRT has utility in discriminating between alternative training systems. That is, using the SRT as a standard post-training measure of effectiveness, the SRT will highlight criterion-based differences between the different CBT systems. In this way, a common standard of comparison (i.e., the SRT) provides an objective aid to identifying the training system likely to yield the most desirable training outcomes. This test and evaluation plan describes a comparison of different training systems using performance on the SRT as a measure of training efficacy.

1.2 Scope

This plan describes the overall CBT system examination strategy and validation criteria to be used in evaluating four candidate CBT programs. The primary measure of training effectiveness for this analysis will be trainee performance on SRT content questions and image tests following training. The programs will be evaluated at Atlanta-Hartsfield International Airport (ATL), Detroit Metropolitan Airport (DTW), and Seattle-Tacoma International Airport (SEA).

1.3 System Description

The four CBT programs to be evaluated in this study are from the ATA, ICTS, SafePassage International, Ltd., and Smart Approach, Ltd.

1.4 Critical Operational Issues and Criteria

The Critical Operational Issues and Criteria (COIC) are those necessary to evaluate the CBT programs. The strategies for evaluating these COICs, and their associated Measures of Performance (MOPs) and Measures of Effectiveness (MOEs), are discussed below.

1.4.1 Issue 1. Absolute Training Effectiveness

Do screeners acquire sufficient knowledge with each CBT program to progress to OJT?

Criterion 1-1. The criterion is investigative in nature.

MOP 1-1-1. SRT scores associated with each CBT system.

MOP 1-1-2. The percentage of trainees who successfully complete each training program.

MOE 1-1-1. Percentage exceeding the 50^{th} percentile score for current screeners.

MOE 1-1-2. A CBT program's Overall Success Measure (OSM) derived as the product of the percentage of screeners who complete training and the percentage of screeners who pass the minimum performance score.

1.4.2 Issue 2. Relative Training Effectiveness

Do the CBT programs differ in their training effectiveness?

Criterion 2-1. The criterion is investigative in nature.

MOP 2-1-1. Post-training differences between CBT programs on the overall SRT score.

MOP 2-1-2. Post-training differences between CBT programs on subsets of SRT content and image questions.

MOE 2-1-1. Profile of differences in average amounts of knowledge screeners have as a function of question and X-ray image subcategories.

2. OPERATIONAL TEST AND EVALUATION

2.1 Test Milestones

This Operational Test and Evaluation (OT&E) will be conducted with screener candidates serving as subjects and provided by different contracted security companies. Thus, the logistics of the testing will require careful coordination and communication between airport security personnel, security company personnel, FAA security personnel, and the Human Factors Engineers (HFEs). Table 1 shows the milestones for planning and reporting the OT&E.

Table 1. Test and Evaluation Milestones

MILESTONE	DATE	RESPONSIBLE ORGANIZATION
Project Plan	April 9, 1999	Veridian/FDC
Test and Evaluation Plan	May 9, 1999	Veridian/FDC
Start Test and Evaluation	September, 1999	Veridian/FDC
Test and Evaluation Report	January 1, 2000	Veridian/FDC

2.2 Test Sites

Each of the three sites, ATL, DTW, and SEA, will be involved in all phases of this study.

2.3 Test Organization

The SRT will be installed on computers where CBT is to be conducted. Each site will be visited twice, lasting approximately seven days per visit. At the first visit, the SRT will be installed on computers at the airport. Security trainers will be trained on how to administer the SRT and in the experimental protocol for this study. Initial data collection may begin at that point, contingent on the availability of screener candidates. This initial visit will be conducted at each of the three airports. After the HFE has left the training facility, the security trainers will continue to collect data until 25 trainees for each of the CBT programs have completed the experimental protocol. When data collection is complete, the HFE will return to the airports to collect the data and debrief the security trainers.

2.4 Subjects

Three hundred security screener candidates, who complete the various CBT systems, are needed for this study. Because some percentage of screeners who begin CBT do not complete training, it is assumed that the number of trainees who initially participate in this study will be somewhat greater.

2.5 Operational Test Procedures

Candidates will be randomly assigned to each level of each condition. If a trainee discontinues training, the security trainer will simply assign the next available trainee to the now vacant subject slot. This will insure that no CBT program lags behind in terms of completing the study due to a higher dropout rate. This will also provide a written record of CBT dropouts for later analysis. The SRT will be used as the measure of screener knowledge and candidates will be tested with the SRT following completion of CBT training.

Data collection will continue until all 300 trainees across the four CBT systems have taken the SRT. Because data collection is contingent on trainee availability, a strict procedural timeline cannot be specified.

2.6 Data Collection

The data for the SRT will be collected and stored in a computerized database. The SRT will automatically store responses for all content and image questions for each trainee who takes the test.

The database will contain screener background information, CBT program identification, and SRT scores for each section. The data for each subject will be transferred into electronic media and stored after testing is completed at each airport site. All of the data will be stored in an Excel 5.0 database.

2.7 Limitations Of The OT&E

A potential limitation on results interpretation involves a possible sampling bias that might be introduced by the CBT programs themselves. Trainees will be randomly assigned to the CBT programs. However, unequal dropout rates could create a bia for one or more of these programs. For example, one CBT program might be harder to complete, resulting in a higher dropout rate. The candidates who do complete the training and subsequently take the SRT might, consequently, have a greater overall aptitude level. The resulting SRT performance showing higher test scores would suggest a better quality of screener-readiness training for that particular CBT program. In reality, however, the differences in SRT performance would be due to a stricter selection criterion which would have weeded out the less able trainees prior to taking the SRT. This possible limitation can be attenuated by use of an OSM which uses both the CBT completion rate and the SRT score to evaluate the various CBT programs.

3. DATA ANALYSIS

3.1 Mixed Design Analysis

Descriptive and inferential statistics will be calculated for all data. A mixed-design analysis of variance will be used to evaluate the SRT differences. CBT Program (four levels) and Test Site (three levels) will serve as the two between-subjects independent variables. Content (six levels)

and Image Questions (five levels) will serve as within-subjects independent variables. This results in a 4 x 3 x 6 x 5 analysis.

3.2 Standardized Metric of SRT Scores

SRT data are being collected from over 600 experienced checkpoint screeners in a separate effort (Fobes & Neiderman, 1999) and will be used to establish expert-level scores. The 50th percentile score for this range is proposed as the cutoff score for trainees. Each CBT program will be evaluated according to the number of its trainees who exceed this score.

3.3 Item Analyses

For the SRT content questions, a descriptive distribution of the accuracy rate for all questions will be
established. The specific knowledge for questions with high accuracy rates and high error rates will be identified.

3.4 Additional Analyses

A high dropout rate for a given CBT program could cause a sampling bias for those trainees from that
program who finally take the SRT. It could also indicate inefficient training materials and/or techniques. For these reasons, an OSM will be calculated for each CBT program taking into account both the percentage of trainees who complete training and the percentage of trainees who attain the minimum performance score. This overall performance measure for each CBT program will be

$$\text{Overall Success} = PC \times PP$$

where PC equals the percentage of candidates that complete training and PP equals the percentage of candidates that pass the SRT criterion score. A low score on this measure, signifying poor CBT performance, could be due to either a low completion rate or a low SRT pass rate.

4. REFERENCES

Fobes, J. L., & Neiderman, E. C. (1999). *Test and Evaluation Plan for Screener Readiness Test validation* (DOT/FAA/AR-99/21). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.

Fobes, J. L., Neiderman, E. C. & Klock, B. A. (1999). *Screener Readiness Test items* (DOT/FAA/AR-99/1). Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.