PB2001-105955

DOT/FAA/AR-01/47

Office of Aviation Research
Washington, DC 20591

Test and Evaluation Plan: X-ray Image Screener Selection Test

Joshua Rubinstein, Ph.D.

Office of Aviation Security Research and Development
Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

April 2001

Test & Evaluation Plan

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the objective of this report. This document does not constitute FAA certification policy.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AR-01/47		

4. Title and Subtitle	5. Report Date
Test and Evaluation Plan: X-Ray Image Screener Selection Test	April 2, 2001
	6. Performing Organization Code AAR-510
7. Author(s) Joshua Rubinstein, Ph.D.	8. Performing Organization Report No. DOT/FAA/AR-01/47
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Federal Aviation Administration Associate Administrator for Civil Aviation Security, ACS-1 800 Independence Ave., S.W. Washington, DC 20591	13. Type of Report and Period Covered Test and Evaluation Plan 14. Sponsoring Agency Code ACS-1

15. Supplementary Notes

This test and evaluation plan was prepared by Terrance M. Nelson, Pamela U. Duke, M.A., and William M. Maguire, PhD., Logicon-Federal Data.

16. Abstract

This document presents a plan to develop and evaluate an X-ray Image Screener Selection Test (XISST). The XISST will be a computer-based job sample selection test, which will predict the effectiveness of X-ray screeners. The selection test, which will involve searching for common objects in X-ray images, will be tested for usability before being fielded. The field test will take place at two airports. Human Factors Engineers will administer the XISST to at least 50 security screeners from these airports who have at least 2 months of Threat Image Projection (TIP) experience. The goal of this evaluation will be to determine the reliability of the XISST and its validity as a predictor of screener TIP performance.

17. Key Words	18. Distribution Statement
X-ray Selection Test Threat Image Projection Reliability Validity	This report is approved for public release and is on file at the William J. Hughes Technical Center, Aviation Security Research and Development Library, Atlantic City International Airport, New Jersey 08405. This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	13	

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

TABLE OF CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

This test and evaluation plan describes the methods that will be used to develop and evaluate the X-ray Image Screener Selection Test (XISST). The purpose of the XISST is to predict the future threat-detection performance of newly hired X-ray operators.

The XISST consists of multiple test items. The test items are X-ray images of common objects stored in baggage with other non-target articles. The task is to determine if target objects (e.g., tools, guns, or flashlights) are present in each bag. The test measures both speed and accuracy of search performance.

The XISST will be usability tested by Human Factors Engineers, and any deficiencies will be identified and corrected. A field test of the XISST will then be conducted using 50 screeners at Seattle-Tacoma International Airport and Reno/Tahoe International Airport. These screeners will have had at least 2 months of Threat Image Projection (TIP) experience. The XISST data will be compared to TIP data to determine the validity of the test as a predictor of X-ray screening performance. The psychometric properties of the test, including reliability, will also be determined from the field data.

ACRONYMS

$\underline{c}$	Response Bias
CTI	Combined Threat Image
DTW	Detroit Metropolitan Wayne County Airport
FAA	Federal Aviation Administration
HFE	Human Factors Engineer
MOP	Measure of Performance
P_d	Probability of Detection
P_{fa}	Probability of False Alarm
RNO	Reno/Tahoe International Airport
SEA	Seattle-Tacoma International Airport
TIP	Threat Image Projection
XISST	X-ray Image Screener Selection Test

1. INTRODUCTION

Federal Aviation Regulation §108.17 requires that X-ray operators undergo initial and recurrent training to ensure the safety of airline passengers and their property. To comply, air carriers procure equipment and train personnel to screen passengers and their carry-on baggage before they board the aircraft. Furthermore, the Aviation Security Improvement Act, Public Law 101-604, mandates that the Federal Aviation Administration (FAA) enhance and improve X-ray baggage screener selection, training, and performance. The Aviation Security Human Factors Program (AAR-510) of the Office of Aviation Security Research and Development is the FAA unit tasked with this responsibility.

An objective of personnel selection is to make accurate staffing decisions and to employ individuals that will be successful on the job. Job sample tests, interviews, applications, and psychological tests are examples of methods used for selection. Accuracy in personnel selection would help to decrease costs that security companies incur by hiring and training individuals who cannot perform at the skill level required of them.

For this project, a job-sample test was developed to identify individuals that may have the necessary skills and abilities to detect potential threat items. A job-sample test is a type of selection test in which actual tasks from the target job are given to applicants to test their abilities. Such a test is advantageous in that it's content is clearly relevant to on-the-job performance, and the strong relationship between the test and the tasks required for the job is evident and fair [1, 2].

1.1 Background

The perceptual and cognitive skills of X-ray screeners are critical to the safety of our nation's airlines. It is in the interest of the aviation security companies to hire the best candidates to perform this complex and important job. The X-ray screener must be able to detect potential threat items (e.g., guns, knives, or explosives) within an X-ray image that may be filled with innocuous clutter. The FAA, therefore, is interested in developing a selection test that will assess the skills and abilities that would be required to perform the task on the job. Earlier research has examined the use of various cognitive-skills tests (e.g., embedded figures and mental rotation tasks) with some success [3]. These selection tests were composed of abstract figures. The goal of the current project is to create a selection test that contains images that are directly related to the images that will be seen on an X-ray monitor.

The FAA has deployed Threat Image Projection (TIP) technology on X-ray machines throughout the country. TIP offers an objective means of measuring an X-ray screener's skills on the job. This means that any selection test developed can be evaluated and its validity established using TIP data. A successful selection test should be a good predictor of TIP performance.

1.2 Purpose

The purpose of this plan is to evaluate an X-ray Image Screener Selection Test (XISST) using X-ray images as test items. The images were developed using TIP technology. The XISST consists

of multiple test items. These items include X-ray images of common objects stored in baggage with other articles. The task is to verify if an item from a specific category (e.g., tools, guns, or flashlights) is present in a specific bag. The test measures both speed and accuracy of search performance.

2. CRITICAL OPERATIONAL ISSUES AND CRITERIA

The critical operational issues and criteria and Measures of Performance (MOPs) that will be investigated in the test and evaluation of the XISST are listed in the following subsections. These issues are essential in evaluating the reliability and validity of the XISST.

2.1 Issue 1 - Validity of the XISST as a Predictor of TIP Performance

Does a screener's performance on the XISST predict TIP performance?

Criterion 1-1. Screener's XISST performance is a significant predictor of TIP performance.

MOP 1-1-1. Screener's Probability of Detection (P_d) on the XISST.

MOP 1-1-2. Screener's Probability of False Alarm (P_{fa}) on the XISST.

MOP 1-1-3. Screener's sensitivity (d′) for the XISST.

MOP 1-1-4. Screener's response bias (c) for the XISST.

MOP 1-1-5. Screener response times for XISST targets/items.

MOP 1-1-6. Screener's P_d for TIP images.

MOP 1-1-7. Screener's P_{fa} for TIP images.

MOP 1-1-8. Screener's d′ for TIP images.

MOP 1-1-9. Screener's c for TIP images.

2.2 Issue 2 - Operator Performance for XISST Categories

How well do screeners perform with different types of images?

Criterion 2-1. Investigative in nature.

MOP 2-1-1. Screener's Probability of Detection (P_d) for each category.

MOP 2-1-2. Screener's Probability of False Alarm (P_{fa}) for each category.

MOP 2-1-3. Screener's sensitivity (d′) for each category.

MOP 2-1-4. Screener's response bias (c) for each category.

MOP 2-1-5. Screener response times for each category.

MOP 2-1-6. Screener's P_d for TIP images.

MOP 2-1-7. Screener's P_{fa} for TIP images.

MOP 2-1-8. Screener's d′ for TIP images.

MOP 2-1-9. Screener's c for TIP images.

MOP 2-1-10. Distribution of test scores for high-clutter images.

MOP 2-1-11. Distribution of test scores for moderate-clutter images.

2.3 Issue 3 - Reliability of the XISST

Do the XISST scores have internal consistency?

Criterion 3-1. Investigative in nature.

MOP 3-1-1. Kuder-Richardson reliability coefficient for inter-item consistency.

2.4 Issue 4 - Usability of the XISST

Does the XISST show good usability?

Criterion 4-1. Investigative in nature.

MOP 4-1-1. Usability evaluation of the XISST interface.

2.5 Issue 5 - Item Analysis

Do XISST items have good psychometric properties?

Criterion 5-1. Investigative in nature.

MOP 5-1-1. Kolmogorov-Smirnov test for normalcy of test distribution.

MOP 5-1-2. Percent of X-ray screeners who answer each item correctly.

MOP 5-1-3. Point biserial correlation of each item with the overall test.

MOP 5-1-4. Distribution of overall test scores.

2.6 Issue 6 - Test Fairness

Is the XISST unfair or biased against specific ethnic groups or genders?

Criterion 6-1. Investigative in nature.

MOP 6-1-1. Background, ethnic, and gender information for X-ray screeners.

MOP 6-1-2. P_d on the XISST for screeners of a specific ethnic group and gender.

MOP 6-1-3. P_{fa} on the XISST for screeners of a specific ethnic group and gender.

MOP 6-1-4. d′ on the XISST for screeners of a specific ethnic group and gender.

MOP 6-1-5. c on the XISST for screeners of a specific ethnic group and gender.

MOP 6-1-6. Response times on the XISST targets/items for screeners of a specific ethnic group and gender.

MOP 6-1-7. P_d of TIP images for screeners of a specific ethnic group and gender.

MOP 6-1-8. P_{fa} of TIP images for screeners of a specific ethnic group and gender.

MOP 6-1-9. d′ of TIP images for screeners of a specific ethnic group and gender.

MOP 6-1-10. c of TIP images for screeners of a specific ethnic group and gender.

3. METHOD

Human Factor Engineers (HFEs) will conduct a usability evaluation and a field test at two airports.

3.1 Test Development

The XISST is a computer-based test with on-screen instructions, sample items, and keyboard responses for all items. It automatically records accuracy and speed of responses.

The creation of the XISST requires developing 180 images from 10 basic level categories. A total of 18 Combined Threat Images (CTIs) per search category were created: 6 containing a search target (e.g., gun) and 12 containing no targets. Half of all CTIs contain high clutter and half contain moderate clutter. Table 1 outlines this design. None of the 10 categories should appear as clutter in any of the CTIs.

TABLE 1. XISST DESIGN

Category	Target CTIs		Non-Target CTIs		Total Per Category
	Moderate Clutter	High Clutter	Moderate Clutter	High Clutter	
Guns	3	3	6	6	18
Pens	3	3	6	6	18
Shoes	3	3	6	6	18
Key Chains	3	3	6	6	18
Tools	3	3	6	6	18
Cell Phones	3	3	6	6	18
Flashlights	3	3	6	6	18
Watches	3	3	6	6	18
Kitchen Utensils	3	3	6	6	18
Knives	3	3	6	6	18
Total	30	30	60	60	180

The X-ray image sets were created at the FAA Aviation Security Laboratory with Rapiscan X-ray machines installed with the TIP system. The FAA provided the test bags to be used. The Screener Readiness Test [4] software was adapted so that it could be used for the XISST image sets. HFEs created instruction pages and a set of data collection requirements that the software engineers integrated into the product. Response procedures are straightforward, and instructions are simple and clear.

3.2 Usability Evaluation

A usability evaluation of the XISST will be conducted prior to the field evaluation. By using the Rapiscan X-ray machines located at the FAA Aviation Security Laboratory, a test sample of approximately 10-20 images will be created. After these images have been created, the XISST will be administered to individuals who had not been involved in the image creation process. HFEs will then evaluate the XISST for level of difficulty (i.e., too easy or too hard), testing speed, accuracy, and image quality. The results of this evaluation will be used in the creation of the full image set.

At the same time, a usability analysis will be conducted on the XISST image software. An analysis on all computer system-related problems will include quality of image test presentation and response interface, accuracy of data storage, and readability of instructions. Any issues that are found to be problematic will be addressed and resolved prior to field testing. The completed test will then be given a full evaluation to determine if it is ready for the field test.

3.3 Field Investigation

3.3.1 System Description

HFEs will administer the XISST on Compaq Presario 1200 and Toshiba Satellite Series 4200 laptop computers with AMD-K6™ 3-D processors, 60 MB of RAM, and the Windows® 98 operating system. The XISST will be presented to participants using Microsoft® Netscape Navigator 4.7. The entire keyboard will be used to collect participants' demographic data. For the actual test, only keys 1 and 2 will be used for the participants to key in responses to the test questions.

3.3.2 Test Sites

The training facilities for security companies at Seattle-Tacoma International Airport (SEA) and Reno/Tahoe International Airport (RNO) will be the sites where the test will be administered.

3.3.3 Participants

HFEs will administer the XISST to 50 screeners who work at the checkpoints at SEA and RNO. The XISST will only be administered to those screeners who have 2 or more months of TIP experience. At least 25 screeners per airport will be needed to complete the data collection for the XISST.

3.3.4 Procedure

The XISST will be installed on four laptop computers for the field evaluation. These computers will be taken to SEA and RNO where the XISST will be administered to TIP-experienced screeners. The goal will be to administer the XISST to 50 screeners across both airports. Each screener will complete the XISST, which will consist of 10 basic level search categories (TIP X-ray images) containing target images for which the screener will be asked to respond with a

"Yes" (they see the target) or a "No" (they do not see the target). All answers will be entered via keyboard and each screener will have 15 seconds per image to respond. The computer will record the screeners' answers, as well as their response latencies. Arrangements will be made to download TIP data from both sites. These data will be used to analyze test validity.

3.3.5 Data Collection

Data elicited from the XISST will be incorporated into an annotated database for analyzing, archiving, and reporting. The database will be developed in Microsoft® Access and will enable the use and modification of the database to be accessed by an independent database designer and/or analyst.

3.3.6 Data Analyses

The XISST and TIP data will be loaded into corresponding databases. Five scores will be derived for each search category: response time, P_d, P_{fa}, d', and c. An overall score will also be computed by averaging all category scores. Additionally, HFEs will compute item scores by averaging item performance across all subjects.

The goal of the data analyses will be to determine the validity of the XISST by comparing screeners' XISST scores to TIP performance. Validity will be examined by looking at correlations of XISST scores (e.g., speed or accuracy) with TIP performance measures. An important question will be whether speed, accuracy, or some derivation of both is the best predictor of performance. The predictive validity of different search categories will be independently evaluated.

Reliability will be examined by using standard measures of inter-item consistency [5]. Item analyses will look at item difficulty, item-test correlations, and scoring distributions. Items that have deficient psychometric properties may be candidates for elimination in a revised test. If the demographic distributions of screeners permit, HFEs will also look for adverse impact of the test and test items on any demographic group. This is the first step in any study of item bias. Based upon these analyses, the quality of the current XISST as a selection instrument can be determined, and any deficiencies in the test can be identified and corrected in a revised test, if necessary.

4. REFERENCES

1. Hunter, J.E., & Schmidt, F.L., "The economic benefits of personnel selection using psychological ability tests," Industrial Relations, 21, 293-308, (1982)

2. Hunter, J.E., & Schmidt, F.L., "Fitting people to jobs: The impact of personnel selection on the national productivity." In M.D. Dunnettte & E.A. Fleishman (Eds.), "Human performance and productivity: Human capability assessment," Earlbaum, Hillsdale, NJ, 1982.

3. Neiderman, E. C., "Test and Evaluation report: Selection Test Battery Evaluation at Seattle-Tacoma International Airport" (DOT/FAA/AR-01/34), William J. Hughes Technical Center, Atlantic City International Airport, NJ, 2000.

4. Neiderman, E., & Fobes, J. L., "Screener Readiness Test Validation" (DOT/FAA/AR-00/01), DOT/FAA William J. Hughes Technical Center, Atlantic City International Airport, NJ, 2000.

5. Nunnally, J. C., & Bernstein, I. H., "Psychometric Theory," Third edition, McGraw-Hill, Inc., NY, 1994.