DOT/FAA/AR-97/63

A Cognitive Model of X-Ray Security Screening: Selection Tests to Identify Applicants Possessing Core Aptitudes

PB98-125735

Eric C. Neiderman, Ph.D.
J. L. Fobes, Ph.D.

Office of Aviation Research

Washington, DC 20591

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

September 1997

Final Report

This document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Rec'
DOT/FAA/AR-97/63		PB98-125735

4. Title and Subtitle	5. Report Date
A COGNITIVE MODEL OF X-RAY SECURITY SCREENING: SELECTION TESTS TO IDENTIFY APPLICANTS POSSESSING CORE APTITUDES	September 1997
	6. Performing Organization Code AAR-510
7. Author(s) Eric C. Neiderman, Ph.D., and J. L. Fobes, Ph.D., AAR-510	8. Performing Organization Report No. DOT/FAA/AR-97/63
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 United States Department of Transportation Federal Aviation Administration Headquarters Associate Administrator of Civil Aviation Security 800 Independence Ave., S.W. Washington, DC 20591	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code ACS-1

15. Supplementary Notes
Prepared by David J. Bryant, Ph.D. and Jean Dunn, NYMA, Inc.

16. Abstract
This report presents a cognitive model of the perceptual and cognitive processes involved in X-ray screening. The model is used to identify 51 psychometric tests that are potentially useful for selection of X-ray screeners. Selection tests are standardized psychometric instruments used to measure specific aptitudes and abilities related to cognitive functioning and job performance. Selection tests may improve the performance of X-ray screeners by ensuring that individuals with the aptitude for success are placed at security checkpoints. The report describes the purpose, procedure, and items of the 51 tests. The tests are evaluated for validity and reliability and against a set of practical considerations. A battery of 14 tests is recommended for operational evaluation and validation as selection tests for X-ray screening.

17. Key Words	18. Distribution Statement
X-ray screener Selection Test Cognitive model aptitude	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	93	

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

Preface

This report presents a cognitive model of perceptual and cognitive processes involved in X-ray screening. The key FAA personnel are J. L. Fobes, Ph.D., Eric C. Neiderman, Ph.D., Steve Cormier, Ph.D., Brenda A. Klock, and J. Michael Barrientos, all employed with the Aviation Security Research and Development Division, Human Factors Program, AAR-510.

Table of Contents

Page

Acronyms ... v

Executive Summary ... vii

1. Introduction ... 1
   1.1 Background ... 1
   1.2 Purpose ... 2
2. The Job of the Airport Security Screener ... 2
   2.1 Screener Training ... 3
3. Cognitive Processes Involved in X-Ray Screening ... 4
   3.1 Task Analysis ... 4
   3.2 Information Processing Approach ... 5
      3.2.1 Overview of Model ... 6
      3.2.2 Cognitive Processes ... 9
4. Selection Tests for X-ray Security Screeners ... 20
   4.1 Selection Tests ... 20
      4.1.1 Speed Tests Versus Power Tests ... 20
      4.1.2 Standardization, Validity, and Reliability ... 20
   4.2 Criteria for Selecting Tests for an X-ray Screener Battery ... 21
   4.3 Perceptual and Cognitive Selection Tests ... 22
      4.3.1 Tests Related to Image Generation ... 28
      4.3.2 Tests Related to Pattern Integration ... 36
      4.3.3 Tests Related to Object Recognition ... 44
      4.3.4 Tests Related to Classification ... 53
      4.3.5 Tests Related to Decision Making ... 58
      4.3.6 Tests Related to Attention and Vigilance ... 61
5. Recommendations ... 64

References ... 68

Appendixes

A - Vividness of Visual Imagery Questionnaire (VVIQ)
B - Spatial Antecedents Questionnaire Items
C - Revised Individual Differences Questionnaire
D - Kirton Adaption-Innovation Inventory

List of Illustrations

Figures Page

1. A Cognitive Model of X-ray Screening....7
2. Outline of Marr's (1982) Theory of Visual Perception....10
3. Example of Items for the Identical Pictures Test....29
4. Example of Items for the Finding As Test....31
5. Example of Items for the Number Comparison Test....32
6. Example of Items for the Maze Tracing Speed Test....32
7. Example of Items for the Choosing a Path Test....34
8. Example of Items for the Map Planning Test....34
9. Example of Items for the Hidden Figure Test....37
10. Example of Items for the Hidden Patterns Test....38
11. Example of Items for the Copying Test....38
12. Example of Items for the Gestalt Completion Test....39
13. Example of Items for the Concealed Words Test....40
14. Example of Items for the Snowy Pictures Test....41
15. Example of Items for the Shape Memory Test....45
16. Example of Items for the Building Memory Test....46
17. Example of Items for the Map Memory Test....46
18. Example of Items for the Card Rotation Test....47
19. Example of Items for the Cube Comparison Test....48
20. Example of Items for the Form Board Test....48
21. Example of Items for the Paper Folding Test....50
22. Example of Items for the Surface Development Test....50
23. Example of Items for the Visual Recognition Test....53
24. Example of Items for the Letter Sets Test....54
25. Example of Items for the Locations Test....55
26. Example of Items for the Figure Classification Test....56
27. Example of Items for the Picture Number Test....56
28. Example of Items for the First and Last Names Test....57
29. Example of Items for the Object Number Test....58

Tables Page

1. Monocular and Binocular Cues....13
2. Descriptive Principles of Visual Object Recognition....14
3. Potential Selection Tests....23
4. Recommended Selection Tests....65

Acronyms

AGARD	Advisory Group for Aerospace Research and Development
AOA	Air Operation Area
APSS/PI	Airline Passenger Security Screener Pre-Employment Inventory
APTS	Automated Performance Test System
ATA	Air Transport Association
BPS	Boredom Proneness Scale
CA-VAT	Computer Administered Visual Attention Test
CNS	central nervous system
CPI	California Psychological Inventory
CSS	Checkpoint Security Supervisor
CTA	Cognitive Task Analysis
ETS	Educational Testing Service
FAA	Federal Aviation Administration
GAO	General Accounting Office
GATB	General Aptitude Test Battery
HFT	Hidden Figures Test
HPT	Hidden Patterns Test
IDQ	Individual Differences Questionnaire
IED	Improvised Explosive Device
JTA	Job Task Analysis
RAS	Reticular Activating System
RBC	Recognition-By-Components
SAQ	Spatial Antecedents Questionnaire
SIC	Screener-In-Charge
STAI	State Trait Anxiety Inventory
STES	Standardized Tests for Research with Environmental Stressors
UTC-PAB	Unified Tri-Service Cognitive Performance Assessment Battery
VSAT	Visual Selective Attention Test
VVIQ	Vividness of Visual Imagery Questionnaire

Executive Summary

The Federal Aviation Administration is responsible for minimizing and ultimately eliminating instances of attack on aircraft. The front line of defense against hijacking and bombing is the airport X-ray screening checkpoint. Thus, there is a continued need to further improve performance of screeners who staff the checkpoints. One way to improve performance is to employ selection tests to assess the aptitude of prospective screener candidates. This will help ensure that individuals with the aptitude for success are placed at checkpoints.

This report presents a theoretical framework to identify relevant selection tests. In particular, selection tests should be chosen in relation to the cognitive processes and strategies of X-ray screening identified in the model. The screening task is complex and involves a range of perceptual and cognitive processes. The cognitive model presented in this report identifies those processes so that tests can be chosen to accurately measure abilities crucial to X-ray screening. The model is developed in the Information Processing Framework. It describes the taking in and manipulation of information that leads to decisions by the screener. The model describes five, broad, largely sequential stages of processing: image generation, pattern integration, object recognition, classification, and decision making. Image generation is the perceptual identification of basic visual features and the visual search for feature conjunctions. Pattern integration is the perceptual organization of visual features and conjunctions to create percepts of coherent forms and objects. Object recognition is the process of matching the visual pattern of an object to internal representations of objects and determining the kind of object corresponding to the image. Classification is the grouping of objects into relevant categories. For the screener, these categories are definite threat, possible threat, and no threat. Decision making is the evaluating of evidence and generating of an appropriate response by using algorithms and heuristics. Two related processes operate on and interact with these stages of processing: attention and vigilance. Attention is the executive control of processing. Attention determines the mental resources devoted to each cognitive process. Vigilance is the maintenance of attention and arousal. The operation of the cognitive processes are discussed in detail.

Selection tests are standardized psychometric instruments used to measure specific aptitudes and abilities related to cognitive functioning and job performance. There are over 2,500 commercially published selection tests that may measure a wide range of human capabilities. The cognitive model provides a framework for identifying potentially useful tests for selection of X-ray screeners. This paper describes 51 selection tests. For each test, the report describes its purpose, procedure, and items. These tests are grouped by the cognitive processes identified in the model of X-ray screening. The report compares the tests to evaluative criteria for a useful test of X-ray screening ability. They must be valid and reliable measures of the X-ray interpretation cognitive process. Practical considerations also determine the usefulness of tests. Tests must be largely visual in nature, easy to computerize, relatively short (i.e., to fit in an overall battery of approximately 1 hr), and commercially available. A battery of tests should feature a mixture of speed and power tests. Each test is reviewed in terms of its procedure and relation to the criteria for a useful selection test.

This report recommends a battery of 14 tests for operational evaluation and validation as selection tests for X-ray screening. The 14 tests meet the preliminary criteria of validity, reliability, and practicality and are likely to prove useful predictors of performance in X-ray screening.

1. Introduction

1.1 Background

Acts of destruction have occurred on air flights around the world. The bombing of Pan Am Flight 103 over Lockerbie Scotland in 1988 serves as an example of the terrible cost in human life and property of terrorist acts (U.S. Congress, 1989). Such acts of terrorism threaten civil aviation and the lives of the public. Consequently, it is the interest of the U.S. Government and the Federal Aviation Administration (FAA) to minimize and ultimately eliminate all instances of attacks on aircraft inside and outside of the U.S.

The FAA assumes primary responsibility for civil aviation security. In cooperation with airports and airlines, the FAA implements security programs. The Air Transportation Act of 1974 requires all passengers and property to be screened by threat detection procedures. Section 315 of the Federal Aviation Act of 1958 directs the FAA to prescribe regulations for the screening of all passengers and property for the presence of unauthorized and dangerous materials. Thus, the FAA supervises security programs for airports to prevent unauthorized persons from entering air operations areas (AOAs), screen passengers and property, and provide ground and in-flight security.

The passenger and carry-on baggage screening program was begun in 1973 to detect potential threats before they enter aircraft or secure airport facilities. The program has been successful in the detection of firearms (FAA, 1991). Passenger and carry-on baggage screening has proven generally good at detecting test weapons submitted by FAA inspectors posing as passengers. There has been, however, a large degree of variability in performance of airport screeners around the country (GAO, 1987). Some sites had very poor success rates. Thus, there is a lack of standardization and effectiveness among many screeners. This performance variability may stem, in part, from an incomplete understanding of the cognitive processes underlying the screening process and, hence, an inability to select individuals with good basic skills to train as airport security screeners.

The threat to civil aviation has changed in the last decade (Fobes, Lofaro, Malone, Fisher, & Berkowitz, 1994). The FAA concentrated on steps to prevent hijacking in the 1980s in response to events such as the hijacking of TWA Flight 847 in the Middle East. In the 1990s, however, there has been a greater availability of sophisticated explosives and means of constructing explosive devices. This has left aircraft vulnerable to bombing. The FAA has increased efforts to provide security measures against prefabricated explosives (e.g., grenades) and Improvised Explosive Devices (IEDs).

The main security measure for explosives detection remains airport X-ray screening. Explosives are generally more difficult to detect through X-ray screening than are weapons such as guns or knives. This is because bombs, especially IEDs, can be largely composed of harmless common objects such as clocks, electronic devices, and plastic or metal containers. Thus, there is a continued need for improvement in operator training and performance in the screening process.

Given the more difficult task of detecting bombs and the tremendous losses associated with acts of terrorism, the checkpoint X-ray screener plays an increasingly more important role in security.

Despite the importance of screener performance, there has been little emphasis on selection of screeners until recently. The Air Transport Association (ATA) developed a questionnaire, the Airline Passenger Security Screener Pre-Employment Inventory (APSS/PI), to assess attributes such as attentiveness, dependability, and attitude assumed to be relevant to screener performance. This questionnaire lacks breadth of assessment, and its validity has not been established. Other tests have been assessed by the FAA, including the Hidden Figures Test (HFT) and the Hidden Patterns Test (HPT). These tests measure visual form perception ability in general. Test scores have been found to correlate with screener performance on a computer-based X-ray interpretation task (Fobes & McAnulty, 1995).

1.2 Purpose

Although few selection tests are used to identify highly qualified personnel for screener positions, selection is important. Standardized selection of security personnel provides a way of hiring a high-quality workforce. Detecting threats and the deterrent provided by the screening process depends primarily on the skill of the personnel who operate the security equipment. Thus, the very best are needed. Selection of high quality screeners may also reduce attrition in the workforce by excluding individuals who are unsuited for screening work.

The issue explored in this report is how to identify selection tests that will be valid predictors of screening performance. To identify potentially useful selection tests, a theoretical framework of the X-ray screener tasks is presented. This framework identifies the cognitive processes involved in X-ray screener performance. These processes can then be targeted for selection. The ultimate goal is to identify or develop selection tests that measure aptitude on the cognitive processes used in screening. The face and predictive validity of these selection tests will then be evaluated. Finally, effort will be directed to developing a battery of valid tests that assess only relevant cognitive processes.

2. The Job of the Airport Security Screener

To understand the cognitive processes involved in baggage screening, one must consider the screener's task. Security checkpoints focus on two major pieces of equipment, the electromagnetometer (metal detector) that passengers walk through and the X-ray baggage screening device that carry-on baggage must pass through. Each device is monitored by one person responsible for the people and items passing through the device. Screeners are expected to take no longer than 10 seconds per bag under normal operational conditions. This report examines baggage screening only, but it is worthwhile to consider other tasks because security personnel cycle through different tasks on 20- to 30-minute schedules, and this has implications for screener performance. Often, there is one security person assigned to visually inspect passengers who pass through the metal detector and another who performs hand-wand inspections of passengers who alarm the metal detector. A Screener-In-Charge (SIC) or Checkpoint Security Supervisor (CSS) is responsible for the entire checkpoint and monitors workload and rotations of personnel. The SIC or CSS does not perform actual screening functions but does handle determinations of questionable items and positive threat detections.

2.1 Screener Training

This report focuses on the X-ray screening of carry-on baggage as the central task. X-ray screening has three broad components. The operator views an X-ray image of a bag, scans the bag for threats or suspicious items, then decides on an appropriate action (pass the bag, call for a hand search, or stop the bag because it contains what appears to be an obvious threat).

Training of baggage screeners is sensitive to the kinds of perceptual and cognitive tasks performed. Screeners are provided explicit training in (a) image analysis and the interpretation of the gray or color tones of X-ray images to identify object densities, (b) recognition of threat items, (c) the use of shape to classify some kinds of threats, and (d) classification of objects into three categories of obvious threat, potential threat, and no threat (EG&G, 1994). The training regimen further recognizes that attention is a vital requirement, and operators are instructed to stop the baggage screening device if attention is drawn elsewhere.

Screeners must be on the look-out for a wide range of threat objects. These include weapons (e.g., guns and knives), hazardous materials (e.g., acids, fireworks, toxins), prefabricated explosives (e.g., grenades), and IEDs. With such a wide range of potential targets, screeners must detect threats on the basis of many different shapes, object densities and, hence, different kinds of X-ray reflection, and different configurations of parts. IEDs, in particular, offer a challenge to screeners. IEDs are generally made from readily available materials and virtually all components except the initiator and explosive are innocent items or parts of innocent items. A review of a publication such as the *Improvised Munitions Black Book* (1981) reveals that many lethal explosive devices can be constructed using everyday objects such as bottles, cardboard tubing, and wristwatches. Even more sophisticated IEDs can be made using travel clocks as timers, commercially available batteries, and tubing to conceal explosives. Thus, few distinctive features can be used to distinguish IEDs from innocent objects. Success in screening for these kinds of explosives will depend on the ability to identify patterns of features related to IEDs.

Baggage screening is often described as a vigilance task (Lofaro et al., 1994a), but it is apparent that screening involves much more. Screeners do not monitor a display for a simple target. Instead, screeners infer objects by extracting patterns of features from X-ray images, integrate patterns to identify coherent objects, recognize and classify the objects, then make decisions regarding the threat potential of the objects. Screening, then, involves a series of sophisticated perceptual and cognitive operations. Therefore, there is a need to select screeners for more than simply motivation or attentiveness.

3. Cognitive Processes Involved in X-Ray Screening

3.1 Task Analysis

To improve security screener performance, a model that specifies screeners' cognitive processes and strategies is presented. This model is based on an analysis of the screener job and its demands. A simple breakdown of the screening task is that screeners (a) view an X-ray image and perceive shapes and forms, (b) scan for threats or recognizable objects, and (c) decide on the appropriate action (e.g., pass the bag, call for a search). This simple analysis identifies a set of cognitive processes necessary to perform the X-ray screening task.

Prior to composing the cognitive model, previous analyses of the cognitive demands of security screening were considered. Kaempf, Klinger, Wolf, and Lofaro (1994) performed a Cognitive Task Analysis (CTA) to collect data on decision making in screening. A range of techniques were used such as think-aloud protocols, simulations, structured and semi-structured interviews, and observation to define the problems faced by screeners. From the CTA, decision-making requirements and cognitive processes involved in screening were identified.

One important finding was that experienced screeners did not examine each bag in detail. Screeners managed workload by adjusting their level of scrutiny to the volume of passengers. Another insight that can be derived from Kaempf et al.'s (1994) finding is that aspects of screening require varying levels of conscious attention, and that attention should be considered as a primary factor in a cognitive model of screening.

Based on observations and interviews, Kaempf et al. (1994) describe screening as a three-step series of judgments. Step 1 is a set of initial perceptual judgments. Screeners search for distinctive features such as dark (i.e., opaque) or odd (i.e., unusual or unfamiliar) visual elements in the X-ray display. Screeners also report evaluating bags "as a whole," judging whether the pattern of features in the bag image may be dangerous. Such reports are vague about the actual cognitive processes; however, they suggest that screeners integrate features of the image to create a perceived pattern of objects, then judge whether that configuration poses a threat.

Screeners rarely attempt to identify all objects in a bag or to scan all objects individually in a serial fashion. Instead, screeners look for forms that stand out as potential threats and require closer inspection. Decisions of which parts of an X-ray image warrant inspection are generally done without conscious awareness. Again, the cognitive processes involved are vague from this analysis. The analysis does suggest an initial automatic parallel search and integration of visual features, followed by conscious attention to specific forms.

Step 2 is performed if an unusual object is encountered. Screeners attempt to recognize individual objects. This is presumably done by comparing the image with internal representations of threat and innocent items stored in memory. Kaempf et al. (1994) describe this as matching the image to a library of stored images acquired through experience. Evaluation of the image as a threat or non-threat can occur rapidly, indicating parallel comparison to stored representations. Identification will be speeded by having a large set of objects stored in memory.

Each memory representation will be associated with at least one visual pattern. Encountering a familiar pattern allows rapid access to an associated object without the need for explicit induction. If only a few items are stored, the screener will have to consciously reason about the image and attempt to infer its identity.

Step 3 consists of classification and decision processes. If the image is matched to a threat object, the screener stops the belt and calls the supervisor. If the image is matched to a non-threat object, the screener determines whether the object is large enough to conceal a threat (i.e., could occlude or contain a threat). If so, the bag is classified as a possible threat, and the screener calls for a bag search. If not, the bag is passed. The determination of size is not always simple because X-ray screening devices scale images to fit the screen and because screeners do not always know the sizes of objects (Lofaro et al., 1994b).

Lofaro et al. (1994a) conducted a Job Task Analysis (JTA) to examine what screeners do when screening for threats. Their analysis confirms and extends that of Kaempf et al. (1994). The JTA based on interviews and observations produced an extensive flowchart that depicts the possible activities and the decision chain of a screener. Categorizing the actions in the flowchart, screeners engage in five kinds of activities. First is image analysis. Screeners look for features that are diagnostic of threats and non-threats such as size, profile or angle, shape/form, opaque shading, and possible bomb components. Second, screeners make an attempt to identify items. This conforms to matching images to stored representations. If an item can be clearly identified, screeners classify the object as threat or non-threat at this point. Third, if the object is unknown or unidentifiable, screeners engage in an explicit reasoning process. Screeners assess whether the item is large enough to conceal a threat or is at an angle that masks its size or features. If either condition is true, screeners conduct an additional search. If not, they pass the bag. Fourth, screeners engage in classification. If an object can be clearly identified, screeners identify the type to which it belongs. The basic categories are threat, non-threat, and possible threat. Threats can be further grouped into weapon, explosive, and hazardous material categories. Fifth, screeners make decisions. Based on the classification of items, the screener selects an action. The screener will stop the bag for an obvious threat, pass a non-threat, or call for a hand search of an unknown item or an item that could be a possible threat.

3.2 Information Processing Approach

The previous studies reviewed provide a good background for understanding screening tasks. Those studies emphasized decision making and did not offer very much insight into the perceptual/cognitive aspects of screening. Also, the studies relied on self-report and observation techniques. Self report and observation assess conscious strategies but do not illuminate automatic and unconscious cognitive processes.

To understand in detail the cognitive processes of X-ray screening and threat detection, an information processing model is needed. The information processing framework is based on the computer metaphor that describes human cognition as similar in function and process to the operation of computers. In particular, the purpose of cognition is to take in information, manipulate it, and perform an adaptive action based on the information. Further, cognition is viewed as being comprised of sets of computational processes known as algorithms and heuristics that determine how information is encoded and transformed. The information processing framework generally ignores issues of the neurophysiology underlying cognition.

Two basic assumptions underlie this kind of model. First, the model assumes stages of processing and an orderly flow of information. Thus, once some stimulus is sensed, information about it is processed by a sequence of more or less discrete processes. The processing at each stage is unique and serves a specific purpose related to the ultimate goal of generating appropriate behavior. Second, information is assumed to be represented in various forms by different processes. Processes transform information by altering the internal mental code that stands for objects, properties, and events in the world. Thus, it is crucial to consider how information is represented at any stage of processing.

Information processing models have definite limitations. These models overemphasize the sequential nature of cognition. Much of cognition employs parallel processing, and it is a simplification to speak of processes occurring in a strict sequence. Also, this approach offers little explanation of the role of environmental factors in cognition. Information processing models tend to present cognition as a closed system of inputs and outputs. Modern research has demonstrated numerous instances in which cognition is modified with respect to environmental factors.

One advantage of the information processing approach is that it provides a framework for clearly describing cognitive operations needed to perform the screening task. This framework illuminates the basic cognitive aptitudes that a person must have to be a screener. Therefore, it is possible to identify appropriate psychological tests to select individuals who have the aptitude to perform the task well.

3.2.1 Overview of Model

Figure 1 presents a diagram of the cognitive model of X-ray screening based on the information processing approach. This model is also based in part on the tasks analyses of Kaempf et al. (1994) and Lofaro et al. (1994a). The model presents more detail concerning the cognitive processes involved, especially at the early, more perceptual stages of processing.

Processing begins with an X-ray image of a passenger bag. The nature of the X-ray image is important because it does not match normal visual input. The X-ray image is two dimensional (2D) but also transparent. That is because X-rays can penetrate non-metallic materials, objects in depth will present overlapping contours and shapes in the picture plane. Unlike traditional video pictures, X-rays fail to convey the three-dimensional (3D) structure of objects. Further, X-ray displays are designed to convey information about the density of objects in the gray or color scale of the image. X-ray image intensity does not indicate the reflective properties of objects as in normal vision.

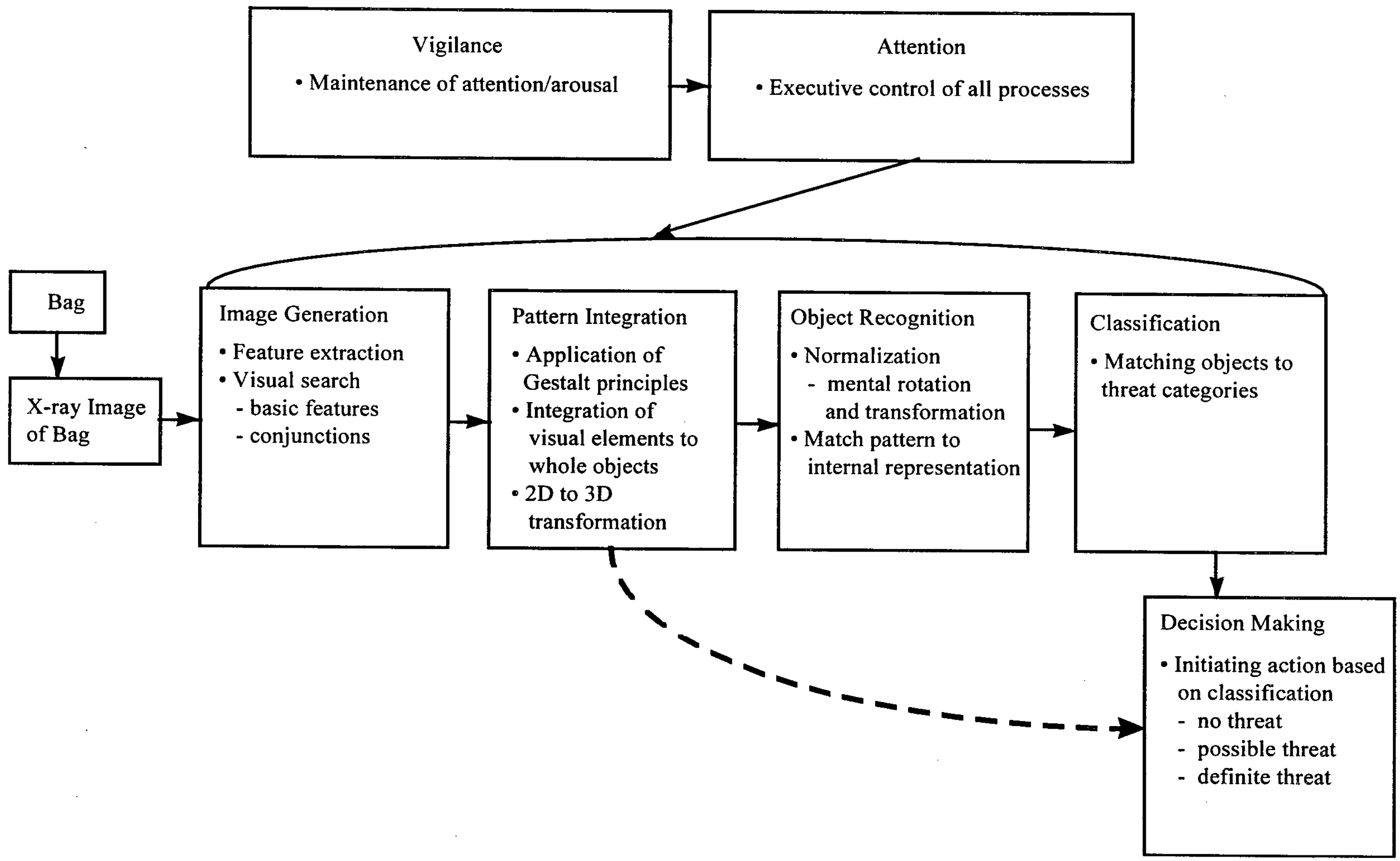


Figure 1. A cognitive model of X-ray screening. Broad stages of processing are indicated in boxes. Arrows depict the normal flow of information, although strict serial processing is not assumed. The dashed arrow from Pattern Integration to Decision Making indicates that screeners may by-pass Object Recognition and Classification. Attention and Vigilance operate as supervisory processes that monitor and influence other perceptual and cognitive activities.

3.2.1.1 Image Generation

Presented with an X-ray image, the first stage of processing is image generation. This stage involves feature extraction and visual search. The goal is to identify visual elements making up the X-ray image. Simple visual elements, such as edges, lines, vertices, and regions of shading, are identified by perceptual mechanisms. Visual search is an automatic process of detecting simple features in a visual background and an effortful process of seeking specific diagnostic features that indicate the presence of a target. It will not always be possible to identify threats or non-threats by the presence or absence of a few basic diagnostic features. More often, the features in a display will have to be integrated to allow identification of objects. Visual search also involves an effortful process of identifying conjunctions of simple features to identify somewhat more complex visual elements.

3.2.1.2 Pattern Integration

After extracting features, pattern integration occurs. In this stage, the viewer employs Gestalt principles of perceptual organization to combine basic features and conjunctions to create percepts of coherent forms and objects. This is not equivalent to recognizing the objects, which requires further processing. The cognitive system identifies where coherent objects are located against the background. Because the X-ray image presents 2D pictures of 3D objects, the viewer will often have to employ transformational processes to infer the 3D structure of forms from their 2D projection. If the patterns formed by objects are clear and unambiguously associated with threats or non-threats, the screener may proceed to the decision stage to determine whether to stop or pass the bag. Screeners reported that they do not always consciously recognize objects in bags but make judgments based on the visual pattern of the X-ray image (Kaempf et. al., 1994).

3.2.1.3 Object Recognition

Once the image has been segmented into discrete forms, the viewer performs object recognition. This is the process of matching a perceived pattern of an object to internal representations of objects and identifying the one kind of object corresponding to the image. Two kinds of processes may be involved. One is viewpoint-independent; the viewer identifies the object's intrinsic axes, codes visual features relative to that frame, then matches that analysis to a structural description of the object in memory. Another is viewpoint-dependent; the viewer normalizes the image to a canonical orientation, typically the viewer's upright, by mental rotation. The normalized image is matched to a representation coded for the viewer's egocentric axes.

3.2.1.4 Classification

Once objects are recognized, the viewer engages in classification in which objects are grouped into categories. For the screening task, only a few categories need be considered. The screener places objects into the threat, non-threat, or possible threat categories.

3.2.1.5 Decision Making

Classification involves decision making. The screener must select an appropriate action based on the classification. The belt should be stopped for an obvious threat, whereas, the bag should be passed for obvious non-threat items. The appropriate action for possible threat items is a more complex decision that will be based on the screener's conservatism, workload, and pressures from passengers and employees to speed the screening process.

3.2.1.6 Attention

Related to all these stages of processing is attention. Attention is the executive control of processing. Thus, attention operates at all stages at all times. The level of attentional resources devoted to any process can vary with the demands of the task. Indeed, some processes, such as feature extraction, are largely automatic and consume few attentional resources. Applying the appropriate cognitive resources at each stage requires a close monitoring of attention.

3.2.1.7 Vigilance

Vigilance is the maintenance and control of attention and arousal. Vigilance is affected by numerous factors including motivation and boredom. If vigilance fails, attention to perceptual/cognitive processes will not be efficient and performance will decline.

3.2.2 Cognitive Processes

This section examines the cognitive processes of security screening in more detail. Subsequent sections examine ways of selecting individuals who are skilled in these cognitive processes.

3.2.2.1 Image Generation

Some basic but crucial visual processing is done by low-level visual systems. Hubel and Wiesel's (1962) classic studies of the cat visual cortex revealed that visual systems are designed to extract basic features. Visual cortical cells respond to specific kinds of light patterns such as edges, bars, and dots. Often, neurons respond to features only at certain orientations. These features are the building blocks of percepts of whole objects.

David Marr (1982) proposed an influential theory of visual perception, which is outlined in Figure 2. Features are detected in or extracted from patterns of light energy to yield a 'primal sketch.' The primal sketch consists of a coding of features at specific positions in the visual field. This is not a representation of objects or even coherent patterns. Rather, the primal sketch represents the features necessary to recognize patterns and objects.

From the undifferentiated field of features in the primal sketch, viewers create a 2½D sketch. This is a representation that identifies surfaces relative to the viewer. Thus, in building the 2½D sketch, the viewer integrates basic visual features to specify edges and planes. Thus, visual features in the primal sketch are converted to a representation of complex features that have defined spatial relations to one another.

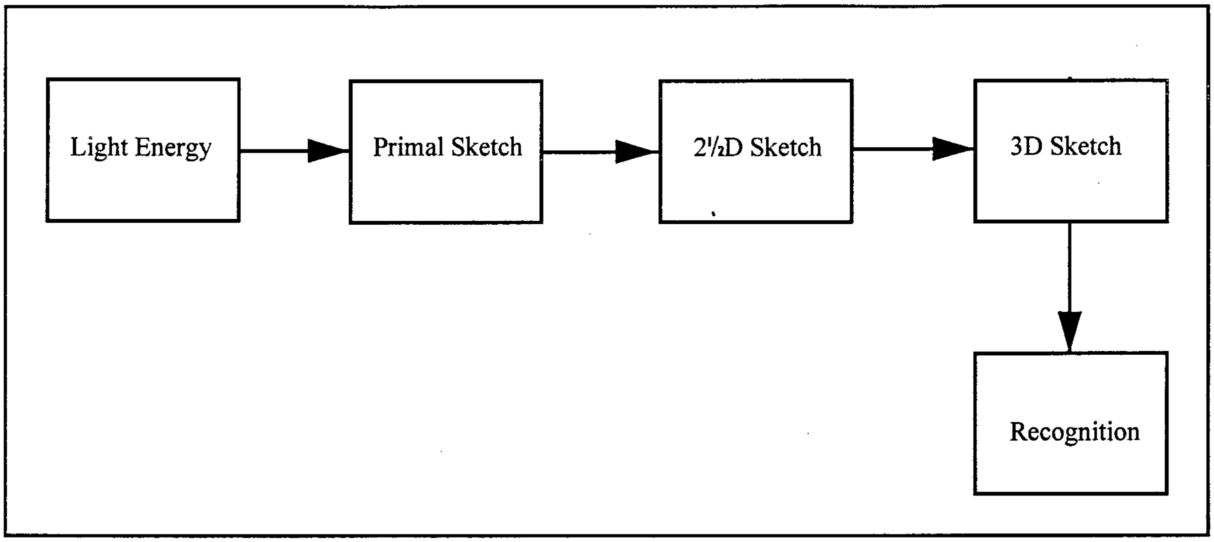


Figure 2. Outline of Marr's (1982) theory of visual perception.

The creation of a 2½D sketch, or perceptual image as it is sometimes called, requires spatial attention. One can process only a portion of the visual field at any one time. Thus, one must select a portion or portions to attend to and arrange the eyes to foveate on that part. Position of the fovea, however, is not equivalent to focus of attention; Posner, Snyder, and Davidson (1980) observed that people can attend to positions as much as 24° from the fovea. Spatial attention is described by the spotlight metaphor. Attention is cast to a region of the visual field like a spotlight illuminating a wall. The spotlight of attention can be shifted to scan different information in the visual field. The spotlight of attention can be adjusted in size so that the region, in degrees of visual angle, of the visual field being processed is relatively small or relatively large. Narrowing the extent of attention allows intense processing of a small amount of information, whereas expanding the extent of attention permits reduced processing of a larger amount of information.

Visual attention is also, in part, feature-based. Treisman and Gelade (1980) proposed a model of visual attention centered on feature extraction and integration. According to this model, the visual system automatically extracts basic features (e.g., lines, angles, colors) in a parallel fashion. Feature extraction corresponds to creating a 2½D sketch. This process requires little conscious effort or attentional resources. It is demonstrated by the pop-out of features from the visual field. When confronted with a field of visual items, single features that differ from the background (e.g., a vertical line in a field of horizontal lines) will be immediately apparent. The time needed to find the target is unaffected by the size of the visual field or the clutter of the background. Identification of complex elements, conjunctions of features, and spatial positions (i.e., creation of the 2½D sketch) requires attention. There is no pop-out of conjunctions of features, even relatively simple conjunctions of simple features. Treisman and Sato (1990) have recently obtained evidence that a small class of special conjunctions may be automatically detected. These conjunctions such as color and depth seem to be important for early visual processing, especially depth perception. Effortfully conjoining features requires attention to be directed to particular features in the visual field and to specific regions. Spatially proximate features can thus be synthesized. Note that depth is normally a crucial factor in establishing degree of proximity. In X-ray displays, depth perception is indirect and may fail.

The attentional demands of feature synthesis depend on the familiarity of patterns and expectations of what patterns will appear (LaBerge, 1973). The greater the familiarity or expectation of a pattern, the easier is synthesis of features and the less attention required. Unfamiliar or unexpected patterns take longer to synthesize and require more attention. There is a greater failure to integrate features appropriately. Practice enhances the speed, accuracy, and automaticity of feature integration and increases familiarity and expectation. Importantly, only small amounts of practice are needed; a few days with a consistent set of stimuli can render feature integration nearly automatic.

3.2.2.2 Pattern Integration

The result of image generation is an organized representation of visual features and conjunctions. From this, the viewer can create a representation of 3D objects that exist in the world and that can be recognized. In Marr's (1982) model, this stage is the creating of a 3D sketch of the visual field.

Feature analysis is one means of integrating patterns based on parallel processing. Objects are defined by a set of visual features. Various objects share features in common, but each object has a unique set of features. The perceptual image can be analyzed for features and conjunctions of features. The presence of features activate representations of objects; that is, they provide diagnostic evidence for the presence of that object. The one object most activated, because all or most of its features are present, is perceived.

The Gestalt Principles of Perceptual Organization (Koffka, 1935; cited in Galotti, 1994, pp. 57-61) describe rules by which the perceptual system constructs a representation of objects by organizing simpler patterns of visual features into coherent forms. There are many Gestalt principles. Among the more important is <u>Figure-Ground Separation</u>, in which edges and shading are used to separate a form from the background of other visual elements. This allows an object to be seen as a definite shape. <u>Closure</u> is the tendency to see forms as complete even when gaps or other distortions disrupt the visual pattern. Viewers perceptually fill obscured or distorted areas. This is useful because viewing conditions are rarely perfect and viewers must compensate for poor images. In <u>Grouping</u>, viewers relate features to one another on the basis of proximity and similarity. Elements that are alike and close are seen as forming a larger pattern or form. Thus, the many simple visual features of the perceptual image are used to identify a larger, more meaningful form. The principle of <u>Maximum Likelihood</u> (Pragnanz) selects the organization of features that yields the most probable shape or form that could give rise to the image. This is based on top-down knowledge of the structure of objects in the world. Pragnanz provides a means for integrating features into patterns that will probably conform to real objects.

Objects in the world are 3D, and the human perceptual system is designed to identify 3D objects. People automatically perceive depth using physiological cues of binocular disparity and binocular convergence (see Sternberg, 1996, pp. 113-116). When viewing a flat picture display, however, the three dimensional structure of things must be inferred entirely from monocular cues. There are many such cues such as texture gradients, relative size, linear perspective, interposition, and aerial perspective (see Table 1). These cues allow a viewer to represent 3D objects and spatial relations from 2D information. This 2D to 3D transformation is central to natural object recognition. Similarly, the ability to mentally transform 2D pictorial information to 3D objects is central to interpreting displays. An X-ray image, however, does not present a full range of depth cues. Therefore, perceiving the 3D forms that give rise to patterns in the X-ray image may be difficult. In particular, the X-ray image lacks texture gradients, aerial perspective, and linear perspective. Interposition is a weak cue because the image is transparent, and objects do not fully occlude one another.

Biederman's (1987) Recognition-By-Components (RBC) theory is a complete model of object recognition, but it specifies pattern integration processes as necessary precursors to recognition. In this model, objects can be decomposed into components, simple geometric volumes such as cones, cylinders, and blocks. These components represent the parts or constituent shapes that make up whole objects. Components can be identified by detecting invariant features (i.e., visual elements that are not dependent on the viewing angle). Basic visual features include collinearity and curvilinearity of points, symmetry, parallelism, and vertices. By detecting invariant features and regions of concavity (e.g., junctions) that specify boundaries or edges, one determines the components present in the visual image. Thus, the cognitive system relates basic features to build a representation of 3D forms.

Sometimes patterns can be identified without an explicit recognition process. Expert chess players, for example, are able to match patterns of pieces on a board to logical moves, whereas, novices fail (Chase & Simon, 1973; de Groot, 1965, 1966). Experts and novices exhibit the same degree of planning ahead, but experts have associated a large number of patterns of pieces with the appropriate move for each. Experts exhibit better memory than novices for chess patterns that could occur in real games.

Performance is based on memory for many patterns encountered over many games played. This requires extensive practice but, with practice, the player learns the correlations between patterns and effective moves. In subsequent games, the expert can frequently bypass explicit recognition of the board configuration and simply recall a move that is associated to the pattern.

A similar process probably occurs in expert X-ray screening. Screeners encounter many common patterns in the displays of bags and learn what outcome (non-threat, possible threat, or threat) is associated with that pattern. The screener can then bypass explicit recognition of objects in the bag and make a decision based on an analysis of the visual patterns (Kaempf et al., 1994). Such correlational learning is most likely for non-threats because these are, by far, the most frequently encountered. With computer-based training (Fobes, Cormier, McAnulty, & Klock, 1996) and Threat Image Projection, it may be possible to increase screeners' familiarity with threats so that they learn correlations of visual patterns associated with threats.

Table 1. Monocular and Binocular Cues

Cues For Depth Perception	Appears Closer	Appears Farther Away
Monocular Depth Cues		
Texture gradients	Larger grains, farther apart	Smaller grains, closer together
Relative size	Bigger	Smaller
Linear perspective	Apparently parallel lines seem to diverge as they move away from the horizon	Apparently parallel lines seem to converge as they approach the horizon
Interposition	Partially obscures other object	Is partially obscured by other object
Aerial perspective	Above the horizon, objects are higher in the picture plane; below the horizon, objects are lower in the picture plane	Above the horizon, objects are lower in the picture plane; below the horizon, objects are higher in the picture plane
Binocular Depth Cues		
Binocular convergence	Eyes feel tug inward toward nose	Eyes relax outward toward ears
Binocular disparity	Large discrepancy between image seen by left eye and image seen by right eye	Small discrepancy between image seen by left eye and image seen by right eye

3.2.2.3 Object Recognition

Object recognition involves the specific identification of items. This process is a joint function of the perceptual representation of discrete 3D forms available through pattern integration and knowledge of objects stored in memory. The viewer needs two things to identify forms: a) well organized and accurate patterns to match to object representations in memory, and b) complete representations in memory of all the kinds of objects the viewer may encounter.

Object recognition is a complex process but, from empirical studies and observation, several things are known about how object recognition operates. These principles are summarized in Table 2 (adapted from Sternberg, 1996).

Table 2. Descriptive Principles of Visual Object Recognition

Principle	**Example**
Recognize familiar things quickly and with a high degree of accuracy	We can easily recognize faces of friends, all the things in our houses, and street signs and other depictions.
Operate on unfamiliar instances	Even though we have never seen a particularly novel example of an object (e.g., a new make of car), we can still identify what kind of thing it is.
Accurately perceive objects that are rotated or observed from different angles	We recognize a coffee cup, even though it is upside down.
Identify objects partly hidden from view, occluded, or in some way noisy	We infer that hidden parts of objects exist, as in the case of the lower torso and legs of TV reporters.
Perform recognition quickly, with subjective ease, and with a large degree of automaticity	We are constantly exposed to changing objects yet process them swiftly and without much effort.

Two classes of theories have been proposed to explain object recognition. The viewpoint-independent approach (Biederman, 1987; Marr, 1982) proposes that recognition proceeds by matching the perceptual image to an object-centered representation. In contrast, the viewpoint-dependent approach (Tarr, 1995; Ullman, 1989) proposes that images are matched to representations defined in an egocentric frame of reference.

In the viewpoint-independent approach, objects are represented by structural descriptions that encode the spatial relations among object components and features relative to the object itself. Viewers extract a description of an object by an object-centered coordinate system during pattern integration and match it to similar object-centered representations in memory (Marr, 1982; Marr & Nishihara, 1978). The matching is done in parallel by comparing the perceptual description to all stored objects and determining the best match. Thus, the perceptual system encodes properties that remain invariant under changes in orientation. Because metric properties (e.g., length, angle) can change dramatically with changes in viewing conditions, researchers (Biederman, 1987; Wagemans, Van Gool, & Lamote, 1996) propose that recognition relies on properties invariant in an affine geometry (e.g., ratio of lengths, aspect ratio). These features can be detected regardless of object orientation, except in a few cases such as a direct top-down view.

Evidence for viewpoint-independent theories comes from studies showing equivalent recognition accuracy and speed across all orientations of alphanumeric and other symbolic characters

(Corballis, Zbrodoff, Shetzer, & Butler, 1978; Corballis & Nagourney, 1978; Eley, 1982). Other evidence comes from studies demonstrating that object naming is facilitated by presentation of partial pictures showing a subset of object components. The amount of facilitation is predicted by the amount of object-centered componential structure available, not the quality of the image in relation to the observer's viewpoint.

Viewpoint-dependent theories propose that object features are encoded in an egocentric reference frame. Objects have a standard or canonic orientation, typically the viewer's upright. During recognition, the image of an object is matched to the standard orientation by a normalization process (Tarr, 1995; Ullman, 1989). The normalization process is typically thought to operate in a fashion similar to mental rotation of imagined shapes (Shepard & Cooper, 1982). Specifically, when an object is viewed, its image is rotated to the nearest familiar viewpoint. This viewpoint may be defined by a single representation for each object, each representing a previously encountered orientation (Tarr & Pinker, 1989). In either case, objects are encoded relative to a viewer-centered frame and normalization is necessary to identify objects.

Evidence for viewpoint-dependent theories comes from studies demonstrating that recognition accuracy and speed vary with the angular difference between the object image and the standard viewpoint. Many studies have documented that naming latencies increase monotonically as objects appear further from the upright (Jolicoeur, 1985, 1988; Jolicoeur & Milliken, 1989; Maki, 1986; McMullen & Jolicoeur, 1992).

The issue of whether there is a single standard orientation or multiple orientations stored in memory has implications for recognition in specialized domains such as X-ray screening. If multiple orientations are stored, new familiar orientations can be acquired through practice, speeding the recognition of misoriented objects. Evidence supports multiple views. Jolicoeur (1985) found that naming latencies increased dramatically with disparity from the upright for the first block of trials, but this effect declined with further trials. Tarr and Pinker (1989) and Tarr (1995) have replicated this effect with 2D and 3D objects and found that naming latency after practice depends on angular distance from the nearest previous view of the object. Thus, people seem to store new viewer-centered representations when they see objects in noncanonical orientations.

Neither viewpoint-dependent nor viewpoint-independent theories can explain the full range of data regarding object recognition (Jolicoeur, 1988; McMullen & Jolicoeur, 1992). Both kinds of information, object-centered invariant features and viewpoint-dependent images, are potentially diagnostic of object identity. Bryant, Reed, and Lanca (1997) obtained evidence of the simultaneous use of viewpoint-dependent and viewpoint-independent processes in visual pattern recognition.

People rarely see objects in isolation, and top-down processing plays an important role in object recognition. One example is the Word Superiority Effect (Reicher, 1969; Wheeler, 1970). Identification of a letter is faster and more accurate when the letter appears in a valid word than when it appears alone or in an invalid word. Similarly, surrounding objects in natural scenes influence recognition of objects (Biederman, Glass, & Stacy, 1973). For example, viewers quickly identify a fire hydrant in a picture of a typical street scene, but are slower and more error prone when the scene is jumbled or the hydrant does not appear in its normal position (e.g., floating in the air). The context limits the number of possible interpretations of objects and aids the matching process.

Top-down processing can be used to recognize objects in X-ray images. Objects do not appear randomly; certain things are typically found in certain places and with certain other kinds of objects. Carry-on baggage typically contains common objects like clothes, toiletries, cameras, and personal electronic devices. The cognitive system is designed to make use of the regularity of objects to aid recognition of individual items.

3.2.2.4 Classification

Upon recognizing objects, the screener classifies them as representing a threat or not. Classification is a complex process. Objects must be grouped into categories that allow simplification, generalization, and prediction. The complexity is somewhat reduced in baggage screening because the categories are fairly well defined. The threat category, for example, is specified by FAA and airline guidelines that indicate what objects are definitely considered threats. IEDs present more of a challenge because they themselves are not standardized. Some components are made from common items, and a screener must judge whether a configuration of objects forms an IED.

The most widely held view of classification is the Prototype Theory (Medin & Smith, 1984; Rosch, 1975; Rosch & Mervis, 1975). Categories are defined by sets of shared features. These are properties that members tend to have. Any member of a category may lack some of these features but will have a significant proportion of features in common with all other members. Thus, a set of central features exists, but any one feature is not necessary for category membership. This theory gives categories two important properties. First, category membership is defined by similarity, not exact rules. Second, membership is fuzzy, such that some instances are perceived to be better examples of the categories than others. Membership is not all or none but a matter of degree. A prototype is a theoretical construct that captures this notion of centrality. It is the best example of the category and is composed of the central features. Instances of the category are related to the prototype by their degree of similarity.

The earlier Classic View held that categories are defined by the presence of critical features (Bruner, Goodnow, & Austin, 1956; Smith & Medin, 1981). All instances must have all the critical features that define the category. Thus, membership is clear-cut and all instances are equally good examples of the category. This view is ideal in X-ray screening: given an object, it is placed unambiguously in the correct category.

The Theory-Based View synthesizes the classical and prototype views (Medin, 1970; Rips, 1989). According to this framework, people use theories or hypotheses about the function of a category to define the dimensions used to assess similarity and assign objects to categories. Further, people form rules that can force inclusion or exclusion of instances regardless of perceived similarity to a category prototype. For example, the presence of a detonator forces an item to be placed into the threat category even if all other components seem like innocent personal items.

3.2.2.5 Decision Making

Screening is an explicit decision-making process and there are many factors that could affect performance. First, people generally exhibit a confirmation bias (Sternberg, 1996, pp. 391-392). This is the tendency to seek evidence that conforms or supports pre-existing beliefs and hypotheses. This bias is problematic because disconfirming evidence is generally more informative. Thus, if items in a bag are questionable, cues that the items cannot be a bomb are helpful in identifying the objects. Doing so leads a screener to evaluate evidence of a threat thoroughly. If a screener seeks to confirm the suspicion of a threat, there is a tendency to look for features consistent with the threat. The screener could miss disconfirming evidence or, worse, distort perception of the X-ray image to identify ambiguous features as evidence of a threat.

Second, people often use heuristics in decision making. Heuristics are rules of thumb or informal procedures that make use of probabilistic information and simplifying assumptions. One example is the availability heuristic (Tversky & Kahneman, 1973). To induce the likelihood of an event, one assesses the ease with which the relevant operation of retrieval, construction, or association can be performed. In other words, if an event is easy to recall or imagine, it is judged to be a likely event. This is not inherently bad; the heuristic is simple and fast, and it works when retrieval is determined by the true probability of events in the world. Retrieval, however, is often determined by a large number of factors, few related to how often something actually occurs. In screening, threats are rare and retrieving instances of threats with particular X-ray images is probably harder than instances of innocent items. If items are ambiguous, screeners may err toward judging the items innocent because they have an easier time retrieving similar patterns that turned out to be non-threats. Every bag is independent, however, so that the probability of a threat in a given bag is not determined by the screener's past experience. Thus, errors of leniency could occur.

The representativeness heuristic is another example of decision making that could affect performance (Kahneman & Tversky, 1973). The subjective likelihood of an instance of some category or type of event is based on how representative of, or similar to, the class is to the instance. In other words, if the instance seems to be a good example of some kind of event, then it is judged to be a likely instance of that event. People overestimate the probability of rare events that have superficial similarities to a category. This could lead to errors of conservatism in screening. If a bag contains wires and metal, a screener may overestimate the probability that the bag contains a threat because these aspects seem representative of a bomb or IED.

It is difficult to know how these heuristics might interact in practice. There is no reason to believe people use heuristics in a consistent manner. Judgment is based on these and other heuristics which can lead to biases in estimating the likelihood of events. This is especially true in situations that do not provide regular feedback on the true rate of events or in situations with extremely rare events.

Another factor that could affect performance is the gambler's fallacy (Galotti, 1994, pp. 346-348). This is the erroneous belief that a random process will keep track of outcomes to ensure that the rate of an outcome in the short run will equal the overall probability of the outcome in the long run. For example, consider observers watching a roulette wheel. They see it come up red on eight successive trials. The true probability of black on the next trial is 50% because each trial is an independent event. The observer, however, will tend to expect a much higher probability of black to balance the run of reds. Screeners are aware that the rate of threats is low, but they could be especially susceptible to the gambler's fallacy if they fail to note that probabilities of bags containing threats are independent. After detecting a threat in one bag, a screener might believe that subsequent bags should not contain any threat to balance the rare occurrence of a threat. This would lead screeners to reduce sensitivity to threats and impair performance.

3.2.2.6 Attention

The study of attention encompasses a broad range of phenomena. The topic most relevant to a model of screener performance is the capacity model of attention. Kahneman (1973) initially proposed the idea that attention can be viewed as a set of cognitive processes for manipulating information. Performance of these operations is limited. Thus, attention is a limited resource for processing information, a general processing capacity.

Attentional resources can be deployed flexibly to various processes and stimuli. A complex stimulus can attract a lot of resources. This means that a large portion of cognitive processes are devoted to that stimulus. Other information may not be effectively processed in this case. Simple stimuli, on the other hand, attract fewer resources and permit many simultaneous processes.

A number of predictions follow from this model. First, performance of one task will decline when competing tasks divert attentional resources. Second, the interference of performance produced by simultaneous tasks is non-specific. Any simultaneous tasks will degrade performance of other tasks if the overall limit of attention is reached. This prediction is not strictly true. There are broad domains of processing in which attention can be deployed somewhat independently, notably verbal and visuo-spatial domains. Thus, verbal processes interfere with visuo-spatial processes to a much smaller degree than other visuo-spatial processes, and vice versa (Brooks, 1967). Third, allocation of attention is flexible and can be altered to suit the demands of incoming stimuli.

Another necessary qualification to the capacity model is that not all cognitive processes require the same degree of attention. A distinction must be made between effortful and automatic processes (Posner & Snyder, 1974). Effortful processes are those that require a high degree of conscious attention. Automatic processes (a) occur without intention and can be elicited by external events, (b) are concealed from consciousness such that a person has little insight into how a process is actually accomplished, and (c) consume few attentional resources. Automaticity is a continuum rather than a dichotomy. Nevertheless, many perceptual and cognitive processes exhibit a high degree of automaticity. For example, elements of visual search and feature extraction, application of Gestalt processes, and object recognition occur without conscious attention and consume relatively few attentional resources. Classification and decision making, however, are more effortful processes.

A number of factors determine the automaticity of processes. Some, like sensory and perceptual processes, are innately automatic. Thus, without training, people perform visual search and pattern integration automatically. This is not to say that even perceptual processes cannot be made more effortful by the demands of the task. If sensory input is limited or degraded, image generation and pattern integration will become more difficult and consume more attentional resources. This is especially relevant to X-ray screening where the image inspected by screeners is unlike normal perception and may contain a cluttered field and ambiguous features. Object recognition is also automatic but depends on extensive knowledge of objects. In a technical domain, people must acquire a mental library of object representations to use in recognition. Again, in X-ray screening, an individual's knowledge of objects, especially threat objects, and their familiarity with the appearance of objects will govern the extent to which recognition requires conscious attention. This is a crucial distinction between novice and expert X-ray screeners.

3.2.2.7 Vigilance

Vigilance has been traditionally defined as sustained attention for the performance of tasks involving the monitoring of displays for targets (Broadbent, 1958). The crucial part of this definition is the notion of sustained attention where cognitive processing must be maintained at a certain criterion level. Vigilance can be seen as a process of maintaining attention for all cognitive processes even those involved in complex tasks. Vigilance is the process that ensures that central nervous system (CNS) arousal remains adequate for the brain to carry out cognitive functions.

Arousal and attention have an underlying biological basis in the Reticular Activating System (RAS). This midbrain region is connected to most areas of the cortex. It sends stimulation to almost all areas of the cortex and affects sensory sensitivity, CNS activation, and activity in the frontal lobes. The frontal lobes, in turn, play a supervisory role in initiating, ordering, and stopping cognitive processes performed by other areas of the cortex (Baddeley, 1990).

The RAS responds to numerous environmental cues, modifying its activity and that of the CNS to the information processing demands of the situation. The less the situation demands, the less arousal produced by the RAS. Thus, people adapt to unstimulating environments with lowered cognitive arousal and attention. The RAS also habituates to stimuli. The repetition or continued presence of a stimulus or situation results in lower CNS arousal.

Consistent with the function of the RAS and CNS, increasing arousal definitely has the effect of improving vigilance. For example, people perform monitoring tasks better if situational factors (e.g., temperature, noise level, feedback) produce arousal at a slightly higher level than that of a person performing an inherently interesting task under no stress (Poulton, 1977). Moderate arousal provides the resources for attention to be applied to basic cognitive processes.

4. Selection Tests for X-ray Security Screeners

4.1 Selection Tests

Psychometric testing uses specified procedures to measure human aptitudes, abilities, and individual differences. There are over 2,500 commercially published psychological tests available. The popularity of psychometric tests stems in large part from the standardization of test procedures for all test takers. Further, psychometric tests provide objective a priori measures of aptitude and ability.

The main purpose of psychological selection tests is to assess aptitudes and abilities relevant to performance of jobs or tasks that have practical importance. The goal is to determine whether an individual is capable of doing a job or at least estimate the likelihood of that person succeeding at the job. For this reason, many psychological tests are used as selection tests to identify prospective employees.

Selection tests are based on several assumptions concerning the measurement of any given ability (Anastasi, 1988). First, it is assumed that the ability is quantifiable and that one can represent degrees of ability by test scores. Second, it is assumed that test items accurately measure the desired ability. Third, it is assumed that people exhibit differing levels of ability. Test scores are assumed to form a normal distribution.

A selection test measures performance on a limited number of tasks or questions. These test items can be collected in a quick, convenient, and inexpensive way. An individual's performance in a more complex real-life situation is predicted from the individual's test score. There are many types of measurement tasks such as questionnaires, perceptual or cognitive problems, and behavior in hypothetical situations. A test taker typically performs a set of one or more types of test items, all targeting a certain ability.

4.1.1 Speed Tests Versus Power Tests

In speed tests, individual differences depend entirely on the speed of performance. All items have a low level of difficulty, but the time limit is so short that no one can finish all items. Performance is measured by how many items a test taker can complete correctly in the time limit. In power tests, the time limit is long enough to allow test takers to attempt all items. The difficulty of items, however, is steeply graded. Some items are too difficult for anyone to solve. Both speed and power tests are designed to prevent perfect scores. This allows the test to distinguish levels of individual differences. In practice, the distinction between speed and power tests is a matter of degree. Tests may emphasize speed or power, but performance is affected by both item difficulty and the time limit of the test.

4.1.2 Standardization, Validity, and Reliability

The three basic features of formal assessment are standardization, validity, and reliability. Standardization refers to the administration of a test to all persons in the same way, under the same conditions. All aspects of the test are crucial, from the specific items, to the time limits for the test, to the way instructions are given. For this reason, computerization is often beneficial, as long as the test taker can understand the instructions as presented by the computer.

Standardization also refers to the establishment of norms or statistical standards for the test. Norms allow the comparison of each individual's score to scores of other test takers in a defined group. This makes the test scores interpretable. The norms indicate whether someone is above or below average and by how much. Norms are most useful when the individuals tested share important qualities such as age, social class, education level, and experience. This limits the number of possible but unmeasured, alternative explanations for differences in test scores.

Validity refers to the degree to which a test measures the psychological ability or trait it was designed to measure. Scores should not reflect the particular way the test was constructed or situational factors. Criterion validity is a way of measuring test validity. Test scores are compared to another, independent measure (criterion) of the ability. The higher the correlation between the test and criterion, the more the test measures the underlying ability.

Construct validity is useful when no ideal criterion exists, which is often the case. It is the degree to which scores on a test correlate with other measures such as other tests, judges' ratings of ability, or experimental results considered valid indicators of the ability being measured. Thus, the theoretical constructs that explain the basis of the ability are compared to the expression of behaviors related to the ability. The scores on the selection test should be consistent with what is known about the ability the test measures.

Face validity is the degree to which test items seem to measure what the test is intended to measure. This is, of course, a subjective assessment. Face validity can be somewhat more objectively measured as the degree to which test takers can describe what the test is to measure. Face validity alone does little to ensure a truly valid and reliable test. It is, however, a useful starting point in evaluating the value of a selection test.

Reliability is the extent to which a test produces consistent scores from test takers over multiple administrations of the test. In other words, it is the degree to which the test elicits the same measure of ability over time. This can be measured as the test-retest reliability, which is the correlation between scores of a group that takes the same test on two separate occasions. The higher the correlation, the greater the reliability of the test. The internal consistency is the degree to which a test yields similar scores across its different parts or different sets of items within the test. This is like the test-retest reliability but only requires test takers to take the test once. Further, there is no concern about the test takers' experiences between two test sessions. Reliability is necessary but not sufficient for a useful test. Validity is paramount to a useful test.

4.2 Criteria for Selecting Tests for an X-ray Screener Battery

Reliability is measured as a correlation coefficient (between test and retest or, more often, between split halves of a test). Correlations usually fall in the .80s or .90s (Anastasi, 1988, p. 115) for widely used tests. Lower coefficients, however, can be statistically reliable. Assessing validity is often harder. One method used is factor analysis, a statistical technique for analyzing interrelationships of behavioral data (Anastasi). Factor analysis measures correlations among tests. Groups of tests having high intercorrelations load on a common psychological factor (i.e., it is possible to infer the presence of an underlying psychological process that determines performance on all the tests). Factorial validity is the correlation of a test with whatever is common to a group of tests or indices of behavior. Correlations between a test and a defined criterion can also be used to assess validity. This measures the association of a test to the psychological factor that governs a group of tests. There are no established guidelines for interpreting validity coefficients. Validity coefficients, however, must logically be lower than reliability coefficients because a test should not correlate more strongly with other tests than itself.

Useful tests have content and face validity. Face validity is somewhat subjective and, more importantly, can only be judged with respect to a model of the ability to be assessed. The model of cognitive processes of X-ray screening developed provides a detailed set of psychological constructs with which to evaluate whether psychometric tests measure abilities, traits, aptitudes, or processes relevant to the screening task. The purpose and procedures of tests are analyzed in this report to determine the extent to which they tap processes specified in the model. Only tests targeting important perceptual and cognitive processes are recommended for further operational assessment.

A number of practical concerns also determines the usefulness of selection tests for wide use in selecting X-ray screeners. First, tests that are primarily visual rather than verbal are more useful. Visual tests generally have simpler instructions and are more straightforward to administer. Moreover, the visual character of tests conveys a degree of content validity to tests because the screening procedure involves a great degree of visual processing. Second, it should be possible to computerize the tests to make them available to all security contractors and to make the tests easy to administer. Computerization also enhances standardization of test instructions, presentation, and scoring. Third, the entire test battery ultimately chosen should be no longer than approximately one hour in length. Therefore, the shorter a test, the better suited it will be for the test battery. Fourth, both speed and power tests should be included in the ultimate test battery. Fifth, it must be possible to obtain permission from the commercial owner to perform studies to determine the validity and reliability of the tests as predictors of screening proficiency and to use the tests in selection of screeners if the tests prove useful.

4.3 Perceptual and Cognitive Selection Tests

This section examines a set of psychometric tests that have been employed to assess a range of perceptual and cognitive abilities. The tests are listed in Table 3 and are ordered in terms of the major processes identified by the model of X-ray screening. Thus, tests are related to image generation, pattern integration, object recognition, classification, decision making, attention, and vigilance. A reference from the primary source, if available, is provided for each test. Information about obtaining each test is also provided.

The list of tests in Table 3 is presented as a set of potentially useful and appropriate tests. Not all of the tests equally meet the criteria for the screener test battery. This section provides a description of each test to elaborate its potential. Detail is provided about the purpose for which

Table 3. Potential Selection Tests

Cognitive Process	**Test (Publisher)**	**Section**	**Features**	**Length**
Image Generation	Identical Pictures (ETS)	4.3.1.1	Computerizable, Visual, Speed/Power	3 min.
	Finding As (ETS)	4.3.1.2	Computerizable w/ special equipment, Visual, Speed/Power	4 min.
	Number Comparison (ETS)	4.3.1.3	Computerizable, Visual, Speed/Power	3 min.
	Maze Tracing Speed (ETS)	4.3.1.4	Computerizable w/ special equipment, Visual, Speed/Power	6 min.
	Choosing a Path (ETS)	4.3.1.5	Computerizable, Visual, Power	14 min.
	Map Planning (ETS)	4.3.1.6	Computerizable, Visual, Power	6 min.
	Simultaneous Pattern Comparison (may not be commercially available)	4.3.1.7	Computerizable, Visual, Speed	1.5 min.
	Reaction Time; 2-choice & 4-choice (may not be commercially available)	4.3.1.8	Computerizable, Visual, Speed	1.5 min.
	Tool Matching (U.S. Dept. of Labor)	4.3.1.9	Computerizable, Visual, Power	5 min.

Table 3. Potential Selection Test (cont.)

Cognitive Process	**Test (Publisher)**	**Section**	**Features**	**Length**
Pattern Integration	Hidden Figures (ETS)	4.3.2.1	Computerizable, Visual, Power	24 min.
	Hidden Patterns (ETS)	4.3.2.2	Computerizable, Visual, Speed	6 min.
	Copying Test (ETS)	4.3.2.3	Computerizable w/ special equipment Visual, Speed	6 min.
	Gestalt Completion (ETS)	4.3.2.4	Computerizable, Visual, Speed/Power	4 min.
	Concealed Words (ETS)	4.3.2.5	Computerizable, Visual/Verbal, Speed	8 min.
	Snowy Pictures (ETS)	4.3.2.6	Computerizable, Visual, Power	6 min.
	VVIQ; Visual Imagery (British Psychological Society)	4.3.2.7	Computerizable, Verbal, Questionnaire	approx. 2 min.
	Rey-Osterreith Complex Figure (Unknown)	4.3.2.8	Not computerizable, Visual, Power	approx. 40 min.
	Wiggly Block Worksample (Williams & Wilkins)	4.3.2.9	Computerizable w/ major revision, Visual, Power	10 min.
	Form Matching (U.S. Dept. of Labor)	4.3.2.10	Computerizable, Visual, Power	Unknown
	Three-Dimensional Space (U.S. Dept. of Labor)	4.3.2.11	Computerizable, Visual, Power	Unknown

Table 3. Potential Selection Test (cont.)

Cognitive Process	Test (Publisher)	Section	Features	Length
Pattern Integration (cont.)	Matrix Analogies Test (Charles E. Merrill)	4.3.2.12	Computerizable, Visual, Power	32 min.
Object Recognition	Shape Memory (ETS)	4.3.3.1	Computerizable, Visual, Speed/Power	8 min.
	Building Memory (ETS)	4.3.3.2	Computerizable, Visual, Speed/Power	8 min.
	Map Memory (ETS)	4.3.3.3	Computerizable, Visual, Speed/Power	12 min.
	Card Rotation (ETS)	4.3.3.4	Computerizable, Visual, Speed	6 min.
	Cube Comparison (ETS)	4.3.3.5	Computerizable, Visual, Speed/Power	6 min.
	Form Board Test (ETS)	4.3.3.6	Computerizable, Visual, Power	16 min.
	Paper Folding Test (ETS)	4.3.3.7	Computerizable, Visual, Speed/Power	6 min.
	Surface Development (ETS)	4.3.3.8	Computerizable, Visual, Power	12 min.
	Matrix Rotation (Naval Health Research Center)	4.3.3.9	Computerizable, Visual, Power	Self-paced
	Spatial Antecedents Questionnaire (not commercially available)	4.3.3.10	Computerizable, Verbal, Questionnaire	approx. 10 min.

Table 3. Potential Selection Test (cont.)

Cognitive Process	Test (Publisher)	Section	Features	Length
Object Recognition (cont.)	Revised Individual Differences Questionnaire (Canadian Psychological Association)	4.3.3.11	Computerizable, Visual, Questionnaire	approx. 8 min.
	Picture Memory; Immediate & Delayed (not commercially available)	4.3.3.12	Computerizable, Visual, Power	5 min. (Immediate) and 30 min. (Delayed)
	Visual Recognition Test (may not be commercially available)	4.3.3.13	Computerizable, Visual, Power	approx. 3 min.
Classification	Letter Sets (ETS)	4.3.4.1	Computerizable, Symbolic, Power	14 min.
	Locations Test (ETS)	4.3.4.2	Computerizable, Visual/Symbolic, Power	12 min.
	Figure Classification (ETS)	4.3.4.3	Computerizable, Visual, Power	16 min.
	Picture Number Test (ETS)	4.3.4.4	Computerizable, Visual, Power	15 min.
	First and Last Names (ETS)	4.3.4.5	Computerizable, Verbal, Power	12 min.
	Object Number Test (ETS)	4.3.4.6	Computerizable, Visual/Symbolic, Power	10 min.

Table 3. Potential Selection Test (cont.)

Cognitive Process	Test (Publisher)	Section	Features	Length
Classification (cont.)	Code Substitution Test (may not be commercially available)	4.3.4.7	Computerizable, Verbal, Power	Unknown
Decision Making	State-Trait Anxiety Inventory (Consulting Psychologists Press)	4.3.5.1	Computerizable, Verbal, Questionnaire	Self-paced
	Responsibility Scale (Consulting Psychologists Press)	4.3.5.2	Computerizable, Verbal, Questionnaire	Self-paced
	Choice Dilemma Questionnaire (Holt, Rinehart & Winston)	4.3.5.3	Computerizable, Verbal, Questionnaire	Unknown
	Kirton Adaption-Innovation Inventory (Mercatus, Inc.)	4.3.5.4	Computerizable, Verbal, Questionnaire	Self-paced
Attention & Vigilance*	AGARD-STRES Reaction Time (Naval Aerospace Medical Laboratory)	4.3.6.1	Computerizable, Visual, Speed	9 min.
	Boredom Proneness (may not be commercially available)	4.3.6.2	Computerizable, Verbal, Questionnaire	approx. 10 min.

Table 3. Potential Selection Test (cont.)

Cognitive Process	Test (Publisher)	Section	Features	Length
Attention & Vigilance* (cont.)	Computer Administered Visual Attention Test (may not be commercially available)	4.3.6.3	Computerizable, Visual, Power	approx. 25 min.
	Stroop Color Word Test (may not be commercially available)	4.3.6.4	Computerizable, Visual/Verbal, Speed	approx. 15 min.
	Beck Depression Inventory (International University Press)	4.3.6.5	Computerizable, Verbal, Questionnaire	Self-paced
	X-ray Image Monitoring Test (to be developed)	4.3.6.6	Computerizable, Visual, Speed/Power	Unknown

- Tests do not assess attention independent of vigilance.

each test was developed, the task performed by test takers, and the relation of the test to the cognitive model of X-ray screening. The test is also compared to the five criteria for inclusion in a test battery, and a recommendation of its potential usefulness is made.

4.3.1 Tests Related to Image Generation

4.3.1.1 Identical Pictures

This test is produced by the Educational Testing Service (ETS) (Ekstrom, French, Harman, & Dermen, 1976). The test was designed to measure perceptual speed, which is the "speed in comparing figures or symbols, scanning to find figures or symbols, or carrying out other very simple tasks involving visual perception" (Ekstrom et al., p. 123). The test is intended to measure skill at extracting features or conjunctions in image generation.

Figure 3 depicts an example of the items used in this test. Items consist of simple visual patterns such as a circle with a dark dot in the center or a triangle with a horizontal line in it. Next to each item are five alternatives, including the test item plus four visually similar distracters. The distracters differ from the test item in the presence of one or more critical features. The test

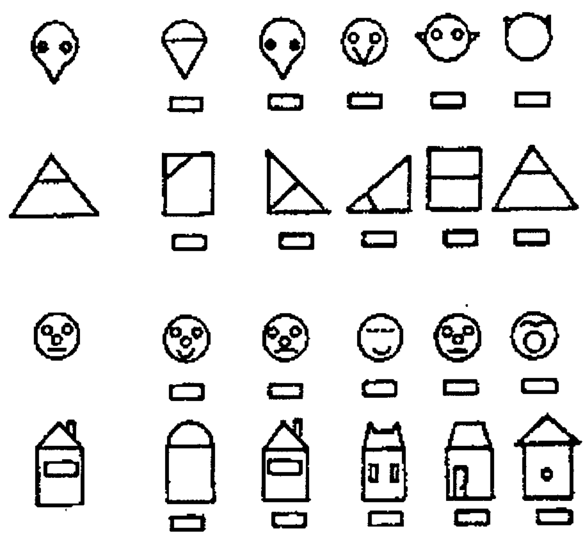


Figure 3. Example of items for the Identical Pictures test.

taker's task is to pick the item from the alternatives as quickly as possible. There are 96 items in 2 parts. Test takers have 1½ minutes to complete each part (i.e., 3 minutes for the entire test). Scores are computed as the number of items correctly marked minus a correction fraction (W/n-1, where W = number wrong and n = number of alternatives) of the number marked incorrectly. The test is nominally a power test, but places a severe time pressure on the test taker.

This test assesses image generation processes. Test takers are asked to quickly extract features of alternatives to identify the one alternative matching the test item on a critical feature or features (e.g., having a dark dot in the center). Items are simple and can be completed with simple feature analysis.

ETS tested the validity and reliability of all tests included in its Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976). All tests generated the appropriate cognitive factor in at least three factor analyses performed in at least two different laboratories. Thus, the Identical Pictures Test has a high degree of construct validity. It also has face validity in the X-ray screener cognitive model. Ekstrom et al. did not report the reliability of the test. Given its repeated administration, however, and inclusion as indicators of cognitive abilities, reliability is likely high. All the tests from the ETS Kit of Factor-Referenced Cognitive Tests have been repeatedly used in diverse settings and have acceptable validity and reliability.

This test could be easily computerized. The test is a short, visual task. The test itself combines speeded and power components, so that it offers a balanced way of assessing image generation. The test is commercially available from the ETS.

4.3.1.2 Finding As Test

This test is produced by the ETS and is designed to measure perceptual speed.

Figure 4 depicts examples of items used in the test. Items consist of words printed in five columns on a page. Test takers quickly scan through the words and cross out only those words containing at least one letter a. There are 200 words in 2 parts, printed 25 words to a page. Test takers have 2 minutes to complete each part (i.e., 4 minutes for the total test). Scores are calculated as the total number of words marked correctly. Difficulty is determined by the time limit.

This test also assesses feature extraction. Test takers identify target words on the basis of the presence of the letter 'a,' which is indicated by the presence of unique features. The features making up letters are simple and test takers do not have time or the need to engage in sophisticated pattern integration or identification of words. The test could easily be computerized, although a new method for collecting responses would have to be developed, perhaps using the mouse to click radio buttons next to words. Like the Identical Pictures Test, the Finding As test emphasizes responding under time pressure. This test, however, uses verbal materials and does not include a broad range of visual features to be detected. For this reason, it is less suited to selection of X-ray screeners than the Identical Pictures Test.

4.3.1.3 Number Comparison

This test is produced by and is commercially available from the ETS and is designed to measure perceptual speed.

Figure 5 depicts examples of items in the test. Items consist of pairs of numbers differing in zero or more digits (e.g., 659 - 659; 73845 - 73855). Test takers inspect items and indicate which pairs contain non-identical numbers. There are a total of 96 items in 2 parts. Test takers have 1½ minutes to complete each part (i.e., 3 minutes for the total test). Scores are calculated as the number of items marked correctly minus the number of items marked incorrectly. Difficulty is determined by the time limit; the test emphasizes responding under time pressure.

This test assesses feature extraction in a fashion similar to the Finding As test, but items require comparison on a variable set of features. That is, test takers must determine the features on which numbers might differ rather than inspecting all items for the same set of features. The test could be computerized. The main drawback of this test is that it is comprised of alphanumeric characters rather than more natural visual patterns. For this reason, it is less suited to selection of X-ray screeners than the Identical Pictures Test; however, it is more useful than the Finding As test because it requires comparison on a variable set of features.

1	2	3	4	5
cider	~~east~~	stripe	insert	defend
bough	blind	coarse	court	settle
fudge	chord	govern	pearl	lodge
greet	~~collar~~	perfect	bridle	oaken
~~fault~~	spoon	special	recess	crown
~~leap~~	piece	consist	soapy	quest
count	rinse	mostly	able	glimpse
shore	drawn	shrink	pledge	every
~~easel~~	fleet	pencil	refuse	break
define	sense	hinder	better	where
entire	uncle	solace	patrol	thorn
ghost	white	keeper	judge	pause
knife	coach	night	defect	hence
hedge	south	clock	trust	short
~~petal~~	period	picnic	other	person
scope	miller	smart	straw	warm
ripen	~~slogan~~	finger	noisy	juice
under	height	useful	defer	enter
~~heard~~	event	slowly	field	ordeal
quite	bond	meant	mend	nurse
jump	west	quick	skill	cool

Figure 4. Example of items for the Finding As test.

659	___	659	7343801	___	7343801
73845	X	73855	18824	___	18824
1624	___	1624	705216831	___	795216831
438	X	436	971	___	971
4321459	___	4814259	446014721	___	446014721
658351	___	656331	5173869	___	5172869
11653	___	11652	6430017	___	6430017
617439428	___	617439428	518198045	___	518168045
1860439	___	1860439	55179	___	55097
90776105	___	90716105	65216067	___	65216057

Figure 5. Example of items for the Number Comparison test.

4.3.1.4 Maze Tracing Speed

This test is produced by and commercially available from the ETS. It was designed to measure spatial scanning ability, which is the speed in visually exploring a wide or complicated spatial field (Ekstrom et al., 1976).

Figure 6 depicts examples of items in the test. Items consist of relatively simple mazes. Test takers are instructed to find a path through the maze and indicate it with a line drawn along the path. The test taker must complete the mazes as quickly as possible and are warned of the need for speed. There are 32 mazes in 2 parts. Test takers have 3 minutes to complete each part (i.e., 6 minutes for the entire test). Scores are computed as the number of mazes through which a path has been correctly drawn. Scoring is somewhat subjective, and scorers are supposed to ignore minor errors such as retracing a path and accidentally crossing a line.

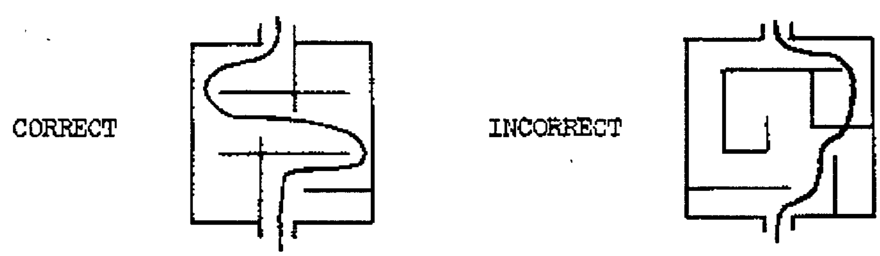


Figure 6. Example of items for the Maze Tracing Speed test.

This test loads on the spatial scanning factor identified by the ETS. This factor involves detecting simple visual features and using them to direct a visual search for other features. Thus, the test measures abilities closely related to image generation. It has been repeatedly administered in ETS test banks and is a reliable test. This test could be computerized but would require specialized input devices (touch screen or drawing pad) to collect responses. It might not be possible to computerize scoring because it requires judgments of the seriousness of errors. The test is short, highly visual, and balanced between speed and power measures.

4.3.1.5 Choosing a Path

This test is produced by and is commercially available from the ETS. It measures spatial scanning ability.

Figure 7 depicts examples of items in the test. Items consist of network drawings depicting multiple lines connecting five pairs of start and finish points. The lines connecting one start point to its corresponding finish point pass through a salient circle. Test takers indicate which of five alternative pairs of start and finish points is connected by a line that goes through the circle. There are 32 visual items in 2 parts. The items progress in difficulty from part one to two. Test takers have 7 minutes per part (i.e., 14 minutes total). Scores are computed as the total number of items correctly marked minus a correction fraction of the number marked incorrectly. This test is more of a power than speed test.

This test is relevant to image generation because the test taker must locate and identify features (e.g., straight paths, corners, junctions) forming a line between the two points. It requires no sophisticated planning or analysis. The test would be fairly easy to computerize; unlike the Maze Speed Test, it requires nothing more than a mouse to indicate responses. It is a relatively difficult test and requires careful instructions.

4.3.1.6 Map Planning

This test is produced by and commercially available from the ETS. It measures spatial scanning ability.

Figure 8 depicts examples of items in the test. Items consist of grid patterns with letters indicating locations around the outer sides. Test takers are instructed that the grid depicts city streets. Circles are located along some line segments. These circles indicate road blocks and prevent tracing a path along that segment. Numbers are printed along line segments inside the grid. These numbers depict buildings that a path must pass by to link two letters. Test takers are instructed to find the shortest path between pairs of letter locations on the map without passing through road blocks. There is always one single shortest path. Test takers indicate their responses to each visual test item by writing the number of the building that is passed to get between the letters. There are 4 grids with 10 paths each in 2 parts. Test takers have 3 minutes to complete each part, for a total of 6 minutes.

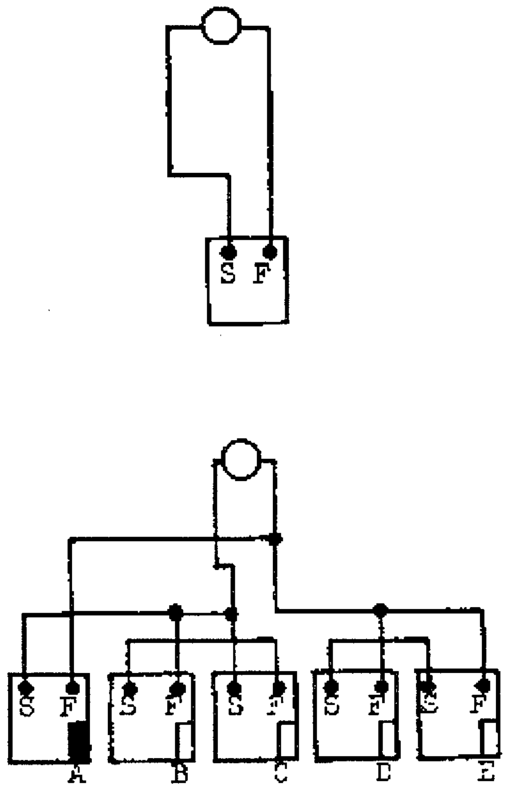
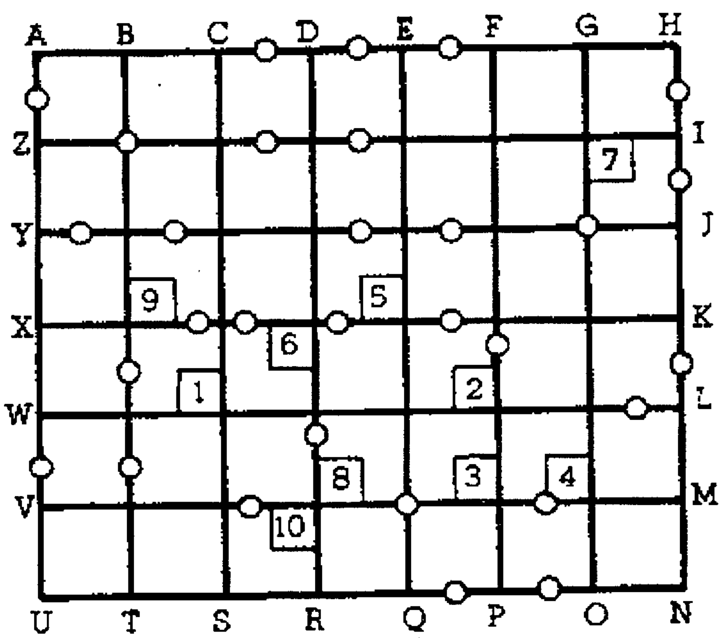


Figure 7. Example of items for the Choosing a Path test.



The shortest route from:	Passes building:
1. A to Z	1
2. E to S	
3. P to J	
4. V to K	
5. O to F	
6. G to M	
7. D to Q	
8. F to T	

Figure 8. Example of items for the Map Planning test.

This test is similar to the Maze Speed Test and requires the same visual scanning and feature extraction abilities. This test, however, could be more easily computerized because test takers can indicate responses with a keyboard and scoring is objective.

4.3.1.7 Simultaneous Pattern Comparison

This test is part of the Automated Performance Test System (APTS) (Kennedy, Wilkes, Dunlap, & Kuntz, 1987; cited in Kennedy, Baltzley, & Turnage, 1989). The APTS is a computerize test battery containing perceptual and cognitive tests. The simultaneous pattern comparison test measures factors relating to target acquisition and visual search. Test takers scan visual patterns and detect featural differences.

Items consist of pairs of eight-dot patterns. These patterns are randomly generated by computer and presented in random order. The number of items is not reported, but the test takes 1½ minutes in total. Test takers judge whether the patterns are similar or different, pressing the 's' key on the keyboard for similar and the 'd' key for different. Scores are computed as the number of pairs correctly identified.

Tests in the APTS have an average reliability of .707 (Kennedy et al., 1989), which is not particularly high but is statistically reliable. A factor analysis generated a validity coefficient of .55 for the simultaneous Pattern Comparison Test with other perceptual tests related to image generation. There may, however, be some overlap with pattern integration processes. The test is already computerized and has been evaluated. It is a short, visual test. The test may not be commercially available.

4.3.1.8 Reaction Time

These tests are part of the APTS. They measure reaction time to visual stimuli.

Items consist of filled and open boxes. In the two-choice version, open boxes appear side by side, above the 'F2' and 'F3' keys on the keyboard. In the four-choice version, four open boxes are presented over the 'F2,' 'F3,' 'F4,' and 'F5' keys. One of the boxes changes from open to filled. Test takers watch the boxes and press the key under the box that changed as quickly as possible. Scores are computed as the total reaction time. The number of items is not reported, but each version of the test takes 1½ minutes.

Tests in the APTS have an average reliability of .707 (Kennedy et al., 1989). A factor analysis generated validity coefficients between .45 and .88 for two-choice and four-choice versions of the Reaction Time Test with other perceptual tests related to image generation. The test assesses an aspect of feature extraction not emphasized by other tests, the speed with which an individual can detect simple visual aspects of a display. The test is computerized and is visual and short. The test may not be commercially available. Reaction Time is a pure speed test.

4.3.1.9 Tool Matching

This test is part of the General Aptitude Test Battery (GATB) (cited in Van de Vijer & Harsveld, 1994). The GATB is a general intelligence speed test that has been under development by the United States Employment Service (a branch of the U.S. Department of Labor) since 1947. The GATB has been used in occupational guidance. Tool Matching is one test used to assess sensory/perceptual fluency.

The test is visual with items consisting of a target figure and four black and white drawings of tools. The test taker indicates which of the four is the same as the target. The response alternatives have the same form as the target but differ in their distribution of black and white parts. There are 49 items and test takers have up to 5 minutes to complete all items. Scores are computed as the number of targets correctly matched.

Van de Vijer and Harsveld (1994) do not report the reliability of this test. A factor analysis revealed a fair association (factor loadings of 1.14 to 1.40) with other tests measuring visual perceptual processes. This test has face validity as a measure of image generation. Responses do not require sophisticated pattern analysis. Rather, test takers attend to differences in simple visual elements, contours, and shading to detect featural similarity. The GATB has been criticized for being excessively speeded (Keesling, 1985). In addition, aptitudes measured by the GATB tend to be intercorrelated, suggesting that individual tests assess multiple, overlapping factors. A computerized version of this test has been created in Dutch (Van de Vijer & Harsveld), which correlates well with the paper-and-pencil version. The English version of the test is available through the Department of Labor, but a computer version would have to be created.

4.3.2 Tests Related to Pattern Integration

4.3.2.1 Hidden Figures

This test is produced by the ETS. It is designed to assess Flexibility of Closure, which is defined as "the ability to hold a given visual percept or configuration in mind so as to disembed it from other well defined perceptual material" (Ekstrom et al., 1976, p. 19). This ability is very closely related to Gestalt principles, notably figure-ground separation. Flexibility of Closure entails perceptually constructing an independent shape from a 2D visual field.

Figure 9 depicts examples of items in the test. Items consist of rectangular figures that have a complex pattern of horizontal, vertical, and diagonal lines drawn within them. Each item contains one, and only one, of five regular geometric target shapes. The target shape is in its upright orientation and the same size as presented for study. Target shapes, however, have line segments crossing it, so that the test taker must identify the target shape from within the overall pattern of lines. There are 32 items in 2 parts. Test takers have 12 minutes to finish each part (i.e., 24 minutes in total). Scores are computed as the number correct minus a correction fraction of the number incorrect.

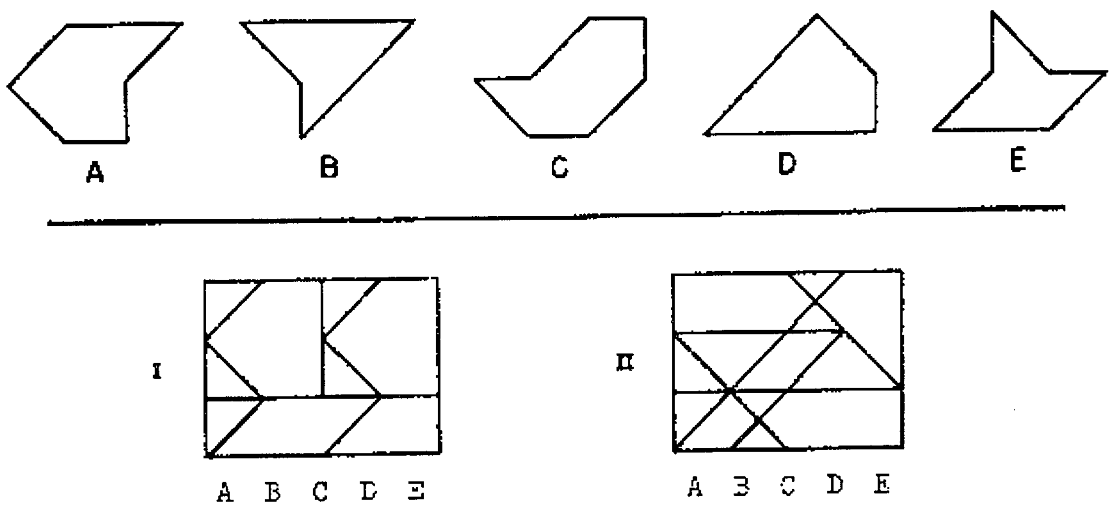


Figure 9. Example of items for the Hidden Figure test.

This test has face validity as a measure of pattern integration because it relies on figure-ground separation and test takers must identify coherent patterns from simpler visual elements. This test could easily be computerized. It is a purely visual test and is commercially available from the ETS. The test is more of a power than speed test, and items are fairly difficult. The test is fairly long, which reduces its practical value.

<u>4.3.2.2 Hidden Patterns</u>

This test is produced by the ETS and is also designed to measure Flexibility of Closure.

Figure 10 depicts examples of items in the test. The test consists of a target pattern (a simple line figure) and more complex patterns in which the target may or may not be imbedded. Test takers view the target then indicate which of the items contain the target in its upright orientation. There are 400 items in 2 parts. Test takers have 3 minutes to complete each part (i.e., 6 minutes total). Each item is simple, and difficulty comes from the time limit of the test. Scores are computed as the number of items correct minus the number incorrect.

This test is similar to the Hidden Figures Test but is more of a pure speed test. It requires perception of a coherent pattern against a more complex background. The test has good validity and reliability. The test is commercially available from the ETS and could be easily computerized.

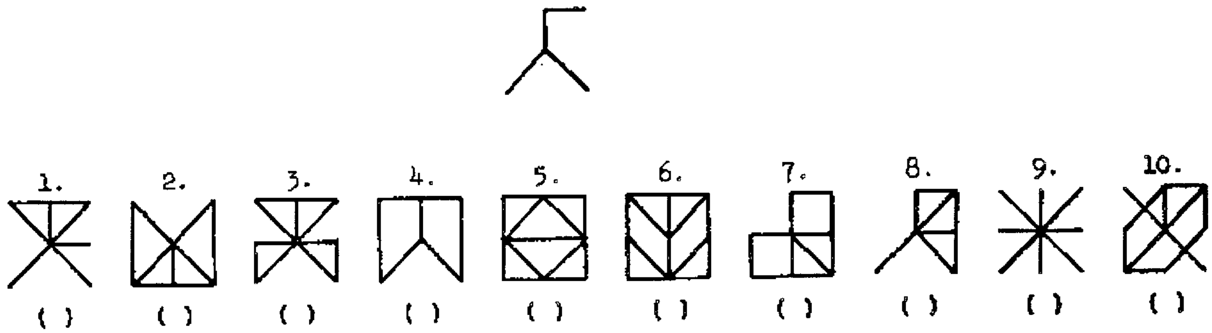


Figure 10. Example of items for the Hidden Patterns test.

4.3.2.3 Copying Test

This test is produced by the ETS and is also designed to measure Flexibility of Closure.

Figure 11 depicts examples of items in the test. Items consist of a five x five grid of dots. One dot is circled as the starting point. Next to the grid is a simple line pattern. One end of the line is circled to indicate the start point. Test takers draw the line pattern onto the grid beginning at the start point. There are 96 items in 2 parts. Test takers have 3 minutes to complete each part (i.e., 6 minutes total). Scores are computed as the number of patterns correctly copied.

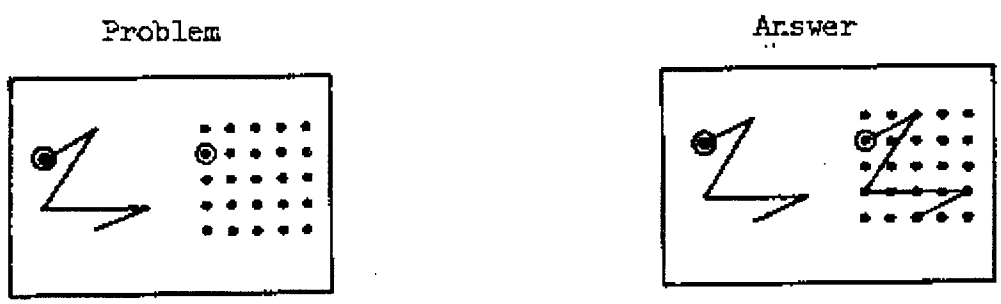


Figure 11. Example of items for the Copying test.

This visual test emphasizes pattern perception and test takers are supposed to keep the pattern in mind to draw it. This test would be difficult to computerize because it would require a special input device (e.g., touch screen or drawing pad). It is primarily a speed test and is commercially available from the ETS.

4.3.2.4 Gestalt Completion

This test is produced by the ETS. It is designed to assess Speed of Closure, which is defined as "the ability to write an apparently disparate perceptual field into a single concept" (Ekstrom et al., 1976, p. 25). The difference between this and Flexibility of Closure is that the viewer sees no obvious closure to start with and does not know what features to seek. In the previous tests, the required configuration was defined beforehand and all the viewer had to do was locate it in a field. Tests of Speed of Closure are very relevant to Pattern Integration because they require the viewer to mentally construct some coherent pattern from ambiguous visual elements.

Figure 12 depicts examples of items in the test. Target items consist of black blotches representing parts of the objects being portrayed. The pictures are similar to highly shadowed objects. Test takers are instructed to use their imagination to fill in the missing parts and name the object in the picture. There are 20 items in 2 parts. Test takers have 2 minutes to finish each part (i.e., total of 4 minutes). Scores are computed as the number of objects correctly identified.

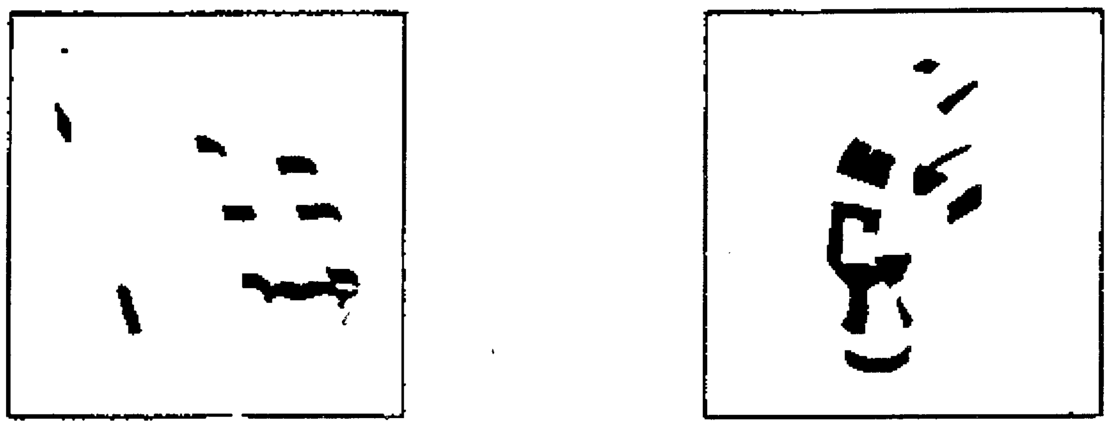


Figure 12. Example of items for the Gestalt Completion test.

This test has face validity as a measure of pattern integration because it forces the test taker to mentally combine elements across the picture to mentally assemble a single coherent form that can be identified. This test could be computerized, but test takers would have to type their answers using the keyboard. This could make scoring difficult because test takers may misspell words. The test is short, purely visual, and is commercially available from the ETS. Difficulty comes from the nature of the items and the time limit.

4.3.2.5 Concealed Words

This test is produced by the ETS and is designed to assess Speed of Closure.

Figure 13 depicts examples of items in the test. Items consist of words printed in a large font that have parts erased. Test takers must identify the words. There are 50 items in 2 parts. Test takers have 4 minutes to complete each part, for a total of 8 minutes. Scores are computed as the total number of items correctly identified.

This test is very similar to the Gestalt Completion Test but uses words rather than pictures. It has somewhat less face validity as an assessment of screening performance because the stimuli are words, which do not emphasize purely visual identification. The test could be easily computerized, although misspellings of responses could be a problem.

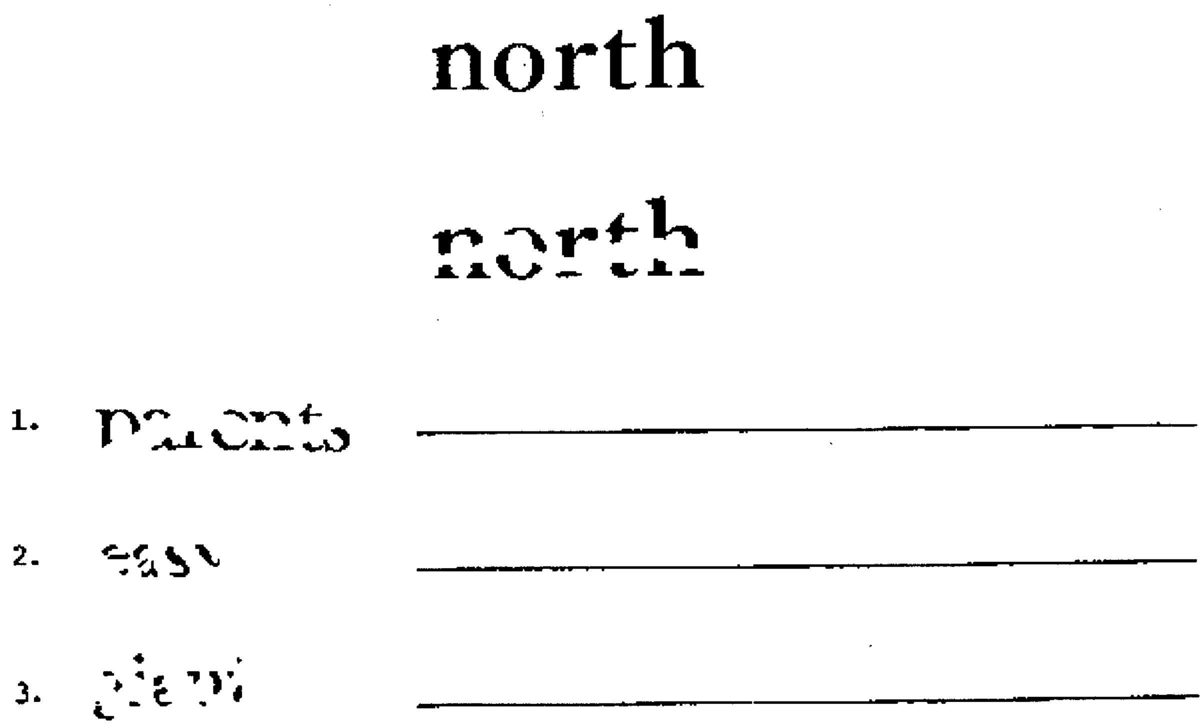


Figure 13. Example of items for the Concealed Words test.

4.3.2.6 Snowy Pictures

This test is produced by the ETS and is designed to assess Speed of Closure.

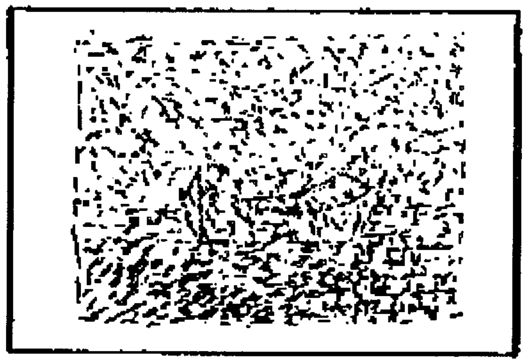
Figure 14 depicts examples of items in the test. Items consist of pictures of objects that are obscured by random small blotches of ink. The overall effect is that of viewing objects through snow or a static pattern on a video display. Test takers view each purely visual item and identify the object in the picture. There are 24 items in 2 parts. Test takers have 3 minutes to complete each part for a total of 6 minutes. Scores are computed as the number of objects correctly identified.

This test has high face validity as a measure of pattern integration. Test takers must mentally integrate disparate elements in the display, pulling together only those features that form the pattern and excluding noise. This test closely matches the task of an X-ray screener who has a degraded image and must identify objects. This test could be computerized, although test takers would have to type their responses and misspellings could be a problem.

4.3.2.7 Vividness of Visual Imagery Questionnaire (VVIQ)

This is a widely available questionnaire that was designed by Marks (1973; cited in Paivio & Harshman, 1983). It was designed to measure the vividness or strength of sensory experience of visual imagery. The experience of imagery predicts how it will be used in perceptual and

1. anchor



2. ______

Figure 14. Example of items for the Snowy Pictures test.

reasoning tasks. If a visual stimulus has a known effect on perception, vivid visual imagery will have a similar effect (Richardson, 1995; Shepard & Cooper, 1982).

The VVIQ elicits self report of imagery experience. Items consist of statements affirming visual elements on a vivid image (e.g., color and shapes of trees in an image of a country scene, color of clothes of family members). Test takers rate the degree of vividness of elements in the statements. There are 16 items, which take approximately 2 minutes to complete. Scores are computed as the total number of points assigned to all items.

This questionnaire correlates with other tests of imagery and with performance on a large number of perceptual and cognitive tasks (Richardson, 1995). No data on the test's validity and reliability are available, but the test is widely used. Visual imagery encompasses mental transformation and pattern integration. Imagery involves mental construction of visual representations and the transformation of those representations. The test is short and could be easily computerized. The questionnaire is entirely verbal. Measurements of imagery, however, usually do not contain visual information to reduce interference of perceptual and imaginal processes. The questionnaire is depicted in Appendix A.

4.3.2.8 Rey-Osterreith Figure

This is a test of pattern perception and memory, which assesses the ability to encode visual patterns and represent them in memory.

The test consists of 1 item: a large, complex visual figure that consists of 18 forms. The components are shapes and line segments. Test takers have 5 minutes to view the figure and copy it on a blank sheet of paper. After some interval (e.g., 40 minutes), test takers attempt to reproduce the figure on another blank sheet of paper. Scoring can be done according to published standards (Taylor, 1959, cited in Annett, 1992). Up to two points are awarded for correctly reproducing a component of the figure. Partial credit is given for incomplete or partially incorrect reproductions.

No data on the validity or reliability of this test are available. The test is relevant to pattern integration because test takers must integrate the components of the figure to store them in memory. This test would be extremely difficult to computerize, and scoring is complex and somewhat subjective. The test takes a long time to administer and may not be commercially available.

4.3.2.9 Wiggly Block Worksample

O'Connor (1928; cited in McDonald & Eliot, 1987) originally developed this test to measure the ability to mentally construct object representations.

Test takers put together wavy-shaped pieces of wood to assemble an oblong block. There are 3 levels of difficulty: 6, 9, and 12 pieces. Daniel (1978) published a test manual with complete scoring procedures, but scores are basically the average assembly time. The test terminates after 10 minutes for test takers who make little progress.

No data on the validity or reliability of the test are available. The test is related to general intelligence (McDonald & Eliot, 1987). It engages abilities needed to perceive relations among parts to form a whole object. The test is highly visual and of moderate length. The test would be difficult to computerize, but it would be possible. A computer version would have to have 3D rendering of the blocks and an interface that allowed test takers to move and rotate the blocks on the computer screen.

4.3.2.10 Form Matching

This test is part of the GATB (cited in Van de Vijer & Harsveld, 1994). Form Matching measures pattern perception and matching abilities.

Items consist of two sets of line drawings of different shapes. For each figure in the first group, the test taker indicates which figure in the second group has the exact same size and shape. There are 60 figures. Sources did not indicate the time limit of the test.

Van de Vijer and Harsveld (1994) do not report the reliability of this test. A factor analysis revealed that the test had factor loadings of 1.09 to 2.36 on a factor related to visual perception. This test taps pattern integration processes, as test takers compare figures and identify identical configurations of features. A computerized version of this test has been created in Dutch (Van de Vijer & Harsveld), which correlates well with the paper and pencil version. The English version of the test is available through the Department of Labor, but a computer version would have to be created.

4.3.2.11 Three-Dimensional Space

This test is part of the GATB (cited in Van de Vijer & Harsveld, 1994). This test measures pattern perception and mental transformation abilities.

Each item consists of a target figure and four drawings of three-dimensional objects. The target figure is depicted as a piece of metal that has to be folded, rolled, or both. Lines in the target figure indicate where the folding should occur. The subject has to indicate which of four drawings of 3D objects can be made of the target figure. There are 40 items. The amount of time per item is not specified.

Van de Vijer and Harsveld (1994) do not report the reliability of this test. A factor analysis revealed that the test had a factor loading of 1.00 on a factor related to visual perception. This visual test taps pattern integration processes as test takers transform perceptual representations to determine the 3D form of objects. The test overlaps with the Paper Folding and Surface Development Tests (see below) and should also measure transformational abilities relevant to object recognition. A computerized version of this test has been created in Dutch (Van de Vijer & Harsveld) that correlates well with the paper and pencil version. The English version of the test is available through the Department of Labor, but a computer version would have to be created.

4.3.2.12 Matrix Analogies

This test was developed by Naglieri (1985a, 1985b; cited in Naglieri & Insko, 1986) to measure nonverbal reasoning ability. It assesses perceptual and cognitive abilities in identifying patterns and inducing relations among visual stimuli.

Items consist of a matrix of three simple visual designs that share some visual properties. The fourth cell of the matrix is left blank. Designs vary in properties such as size, shape, color, and direction. Each item has a set of six test options, one of which shares the properties of the matrix items and fits the fourth cell. The extended form consists of 64 items organized into four groups, the Pattern Completion, Spatial Visualization, Reasoning by Analogy, and Serial Reasoning. Items in the Pattern Completion group require test takers to examine the directions and shapes in an item to determine which of six options accurately completes the pattern. Spatial Visualization items are solved by combining the diagrams within a row and column to obtain the solution or by imagining how a shape will look when it is manipulated. Reasoning by Analogies items require the test taker to analyze the matrix diagram on the basis of specific features (e.g., shape, size, shading) that change and determine how these changes converge to result in a new figure. Serial

Reasoning items require test takers to discover the order in which features appear in the matrix diagram to decide which option completes the matrix. Raw scores for both Item Groups and Total Test can be converted to standard scores for comparison with norms.

The Matrix Analogies Test has a median reliability coefficient of .93 (Naglieri, 1985b). Test-retest reliability is somewhat lower (median = .75). The test possesses a degree of validity. A factor analysis has revealed that the Pattern Completion and Spatial Visualization items load on the same factor, which is related to pattern integration; 83% of items have factor loadings of greater than .40. Thus, the test is not a pure measure of pattern integration, but clearly taps visual perceptual ability. The test does not distinguish well in the upper age ranges among test takers of superior ability or in the lower age ranges among test takers of inferior ability. Items have no time limit, but test takers should require perhaps 20 to 30 seconds per item (i.e., a total of 32 minutes). A shorter version exists (about half as long). The test is exclusively visual and would be relatively easy to computerize. It is primarily a power test and is available through Charles E. Merrill publishers.

4.3.3 Tests Related to Object Recognition

4.3.3.1 Shape Memory

This test is produced by and is commercially available from the ETS. It is designed to measure Visual Memory, which is defined as "the ability to remember configuration, location, and orientation of figural material" (Ekstrom et al., 1976, p. 109). This ability relates to object recognition, which relies on a stored set of visual objects.

Figure 15 depicts examples of items in the test. Items consist of abstract patterns of irregular shapes. The shapes vary in size and shading that are either filled or empty. After viewing a pattern, test takers are presented with test patterns and judge whether the test pattern was part of the studied pattern. This judgment depends on memory for shape and spatial relations between shapes. There are two study patterns, which are page-sized. Each study pattern has 16 test patterns, which are approximately 1x1 inch. Test takers have 4 minutes to view each study pattern and 4 minutes to judge the associated test patterns. Scores are computed as the number of items correct minus the number incorrect.

This test assesses memory for visual patterns, which is a component of object recognition. The test, however, uses simple, noncoherent shapes, which are unlike real objects. The test could be easily computerized. It is of moderate length and a purely visual memory test (test takers could not rely on verbal re-coding for the abstract forms). The difficulty of the test is a joint function of the items and the time limit.

4.3.3.2 Building Memory

This test is produced by and is commercially available from the ETS. It measures Visual Memory.

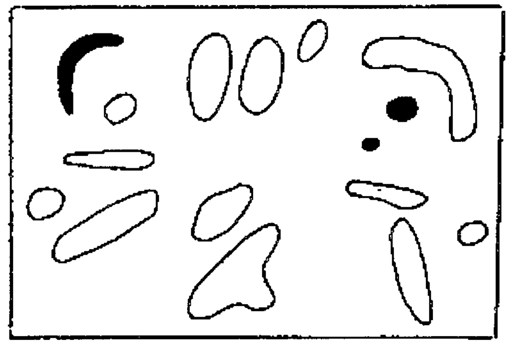


Figure 15. Example of items for the Shape Memory test.

Figure 16 depicts examples of items in the test. Items consist of street maps showing roads, railway tracks, water, and other features. Buildings are drawn at various locations. The buildings are visually distinctive and clearly placed on individual lots. Test takers study the map to learn where buildings are located. In the test phase, the map is shown with the buildings removed. Locations on the map are indicated by letters on the map, and the building symbols are shown next to the map. Test takers indicate the location of each building by writing the letter corresponding to the correct location next to the building. There are 2 maps with 12 buildings each. Test takers have 4 minutes to study each map and 4 minutes to locate the buildings in the test phase (i.e., a total of 16 minutes). Scores are computed as the number of buildings correctly located minus a correction fraction of the buildings located incorrectly.

This test employs meaningful stimuli and is closer to recognition in natural settings than the Shape Memory Test. Maps, however, are not exactly like physical objects. The test could be easily computerized. It is a purely visual memory test (test takers could not rely on verbal re-coding for the abstract forms). The difficulty of the test is a joint function of the items and the time limit.

4.3.3.3 Map Memory

This test is produced by the ETS and is designed to measure Visual Memory.

Figure 17 depicts examples of items in the test consisting of aerial maps of city streets on which buildings are shown as dark shapes. Test takers study a set of 12 maps then pick the studied maps from a set of 12 test maps. Only five of the study maps are shown in the test. The other test items are visually similar alternatives. There are 2 parts to the test, each comprised of 24 maps with 12 test items each. Test takers have 3 minutes to study each set of 12 maps and 3 minutes to indicate memory judgments for total of 12 minutes. Scores are computed as the number of test items correct minus the number incorrect.

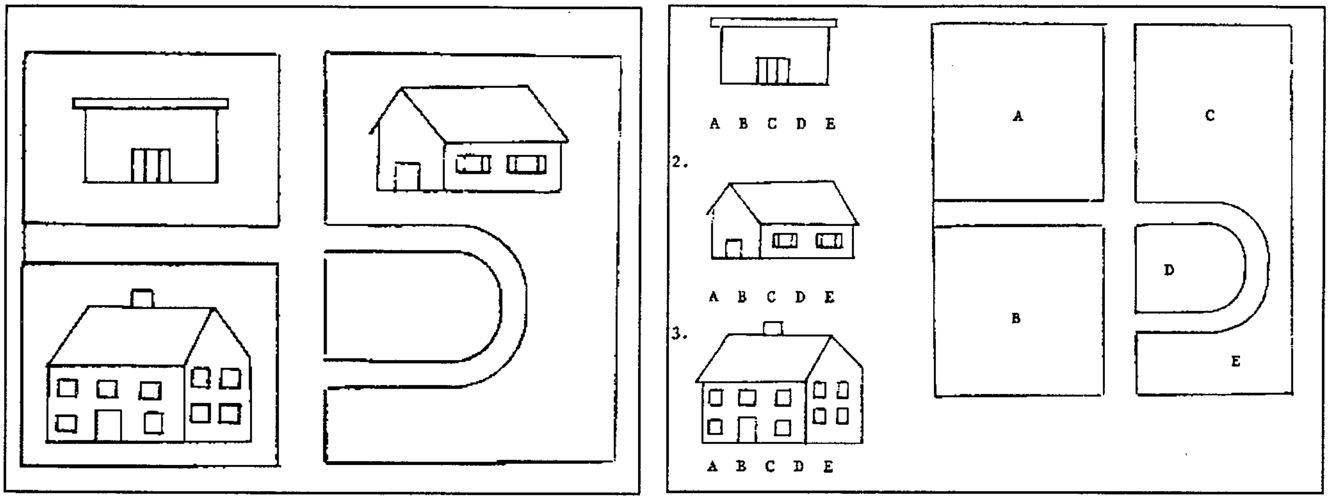


Figure 16. Example of items for the Building Memory test.

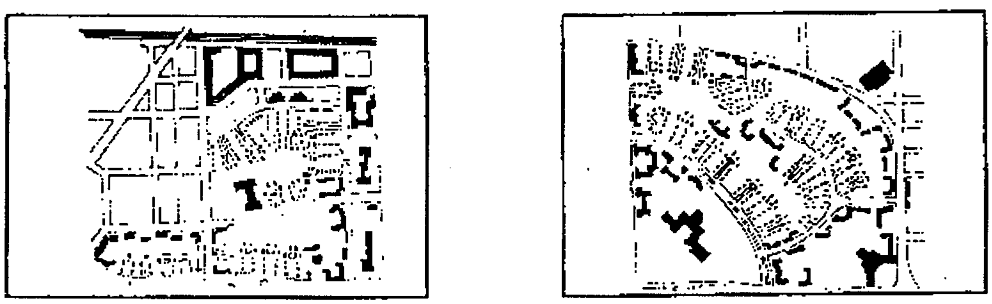


Figure 17. Example of items for the Map Memory test.

The stimuli are maps but have no inherent meaning. The maps form patterns unlike natural objects, but the test does assess memory for patterns and spatial relations. The test could be easily computerized. The difficulty of the test is a joint function of the items and the time limit. The test is commercially available from the ETS.

4.3.3.4 Card Rotation

This test is produced by and is commercially available from the ETS. It was designed to measure Spatial Orientation ability, which is defined as "the ability to perceive spatial patterns or to maintain orientation with respect to objects in space" (Ekstrom et al., 1976, p. 149). Essentially, this refers to the ability to mentally rotate or transform objects in the 2D picture plane or in 3D space. This ability is highly relevant to object recognition, which depends on mental transformation processes to normalize perceptual images prior to recognition.

Figure 18 depicts examples of items in the test. Items consist of simple visual patterns. The patterns are all asymmetric so that distinctive mirror-reversed stimuli can be produced as

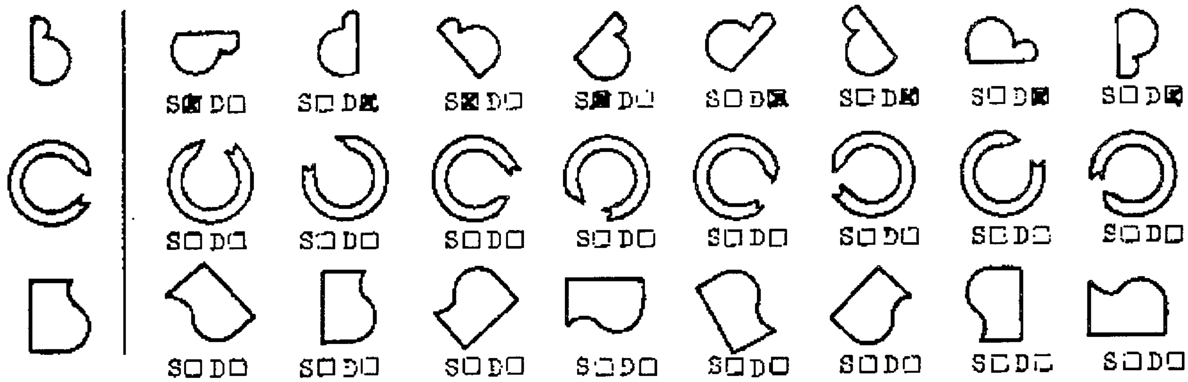


Figure 18. Example of items for the Card Rotation test.

distracter items. Stimuli appear in the left column followed by eight test items. The test items are the original pattern or a mirror-reversed distracter. The test items appear in orientations from 0° to 270° in 45° increments. Test takers indicate for each test item whether it is the same or different from the target pattern. There are 20 target patterns with 8 test items each in 2 parts. Test takers have 3 minutes to complete each part, for a total of 6 minutes. Scores are computed as the number of test items correct minus the number incorrect. The items in this test are generally easy, so the test is primarily a speed test.

The test directly measures mental rotation ability, which is a component process of object recognition. Being able to mentally transform objects is important in X-ray screening because items in a bag may be viewed from any orientation. This test could be easily computerized.

4.3.3.5 Cube Comparison

This test is produced by the ETS and measures Spatial Orientation ability. It requires the 3D transformation of images.

Figure 19 depicts examples for items in the test. Items consist of perspective drawings of 3D cubes. The cubes have a letter on each side, like a child's wooden block. Three sides of each cube are visible. Two cubes are shown side-by-side. Test takers indicate whether the cubes could be the same or are different. Test takers compare the letters on the sides of the cubes to determine whether one cube could be rotated to show the letters in the positions of the second block. The letters must be in the appropriate orientation after the cube has been rotated. There are 42 items in 2 parts. Test takers have 3 minutes to complete each part, for a total of 6 minutes. Scores are computed as the number of items correct minus the number incorrect.

The test measures mental rotation ability in 3D space. It is somewhat more directly related to the X-ray screening task than is the Card Rotation Test because it requires the test taker to transform a 2D image to a 3D representation of an object and to rotate the 3D representation. Screeners must similarly induce the 3D structure of objects from 2D patterns. This test could be easily computerized. It is a short visual test and is commercially available from the ETS. The items in this test are generally easy, so the test is primarily a speed test.

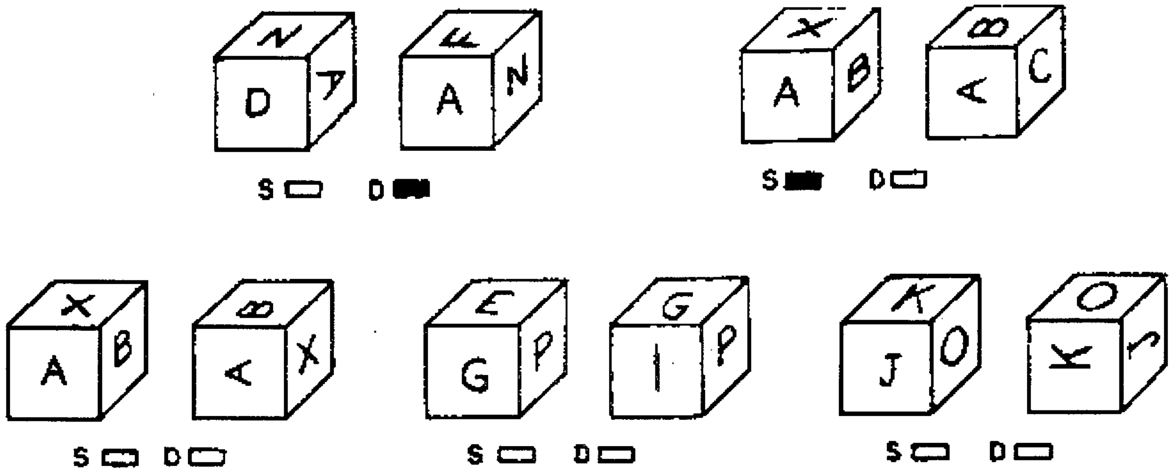


Figure 19. Example of items for the Cube Comparison test.

4.3.3.6 Form Board Test

This test is produced by and is commercially available from the ETS. It was designed to measure Visualization, which is defined as "the ability to manipulate or transform the image of spatial patterns into other arrangements" (Ekstrom et al., 1976, p. 173). This ability is related to processes of converting 2D to 3D forms as an aid to recognition. Visualization is similar to mental rotation but emphasizes additional mental transformations to infer the complete structure of objects.

Figure 20 depicts examples for test items consisting of a geometrical figure. Under each figure are rows containing shaded pieces. Test takers mark only those pieces that can be arranged to make the complete geometric figure. Any number of shaded pieces can be used in the correct answer, and a piece may be used more than once. There are 12 figures with 6 rows of pieces in each of 2 parts. Test takers have 8 minutes to complete each part (i.e., 16 minutes total). In the paper and pencil version, test takers are also allowed to draw sketches to aid solution. Scores are computed as the number of pieces correct minus the number incorrect.

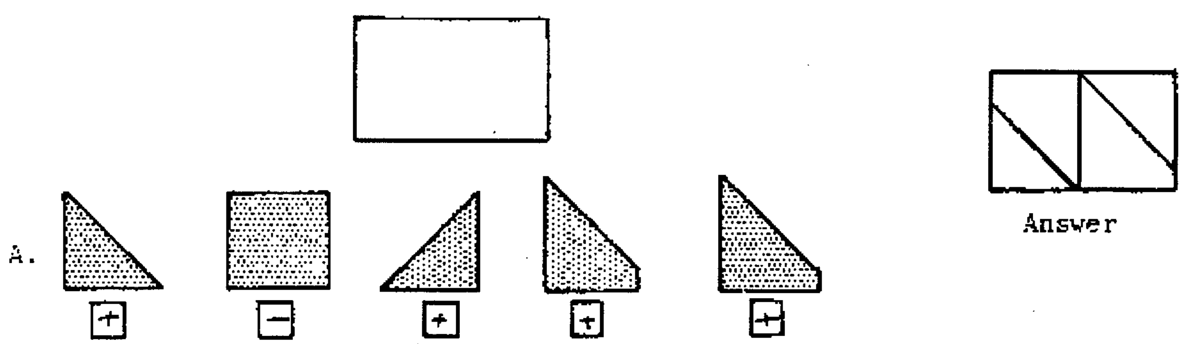


Figure 20. Example of items for the Form Board test.

This test requires test takers to hold in mind several shapes and mentally transform their arrangement. It is a general test of visualization but limited to two dimensional items. For this reason, it does not map onto X-ray screening tasks as closely as other tests. The test could be easily computerized. The difficulty of items is derived mostly from the items themselves rather than the time limit of the test.

4.3.3.7 Paper Folding Test

This test is produced by the ETS and is designed to measure Visualization.

Figure 21 depicts examples for test items consisting of drawings of folded paper. A target, presented to the left, shows a piece of folded paper with a hole punched through all the layers of paper. Test items, presented to the right, show unfolded sheets of paper with multiple holes. Test takers indicate which of the test items is consistent with the target. That is, they select the one sheet that shows where the holes will be when the target is unfolded. There are five test items paired with each target. There are 20 target items in 2 parts. Test takers have 3 minutes to complete each part (i.e., a total of 6 minutes). Scores are computed as the number of items correct minus a correction fraction of the number incorrect.

This test requires test takers to imagine changes in an object. Thus, it taps processes of mental rotation and transformation of object representations. Further, it requires 3D visualization. The test could be easily computerized, and it is commercially available from the ETS. It is a short, purely visual test, which is a joint speed and power test.

4.3.3.8 Surface Development

This test is produced by the ETS and is designed to measure Visualization.

Figure 22 depicts examples for test items consisting of pairs of drawings. One depicts a piece of paper that can be folded on dotted lines (shown in drawing) to form the 3D object shown in the other drawing. Each picture has an 'x' marking one side. The drawing of the unfolded paper has numbers 1 to 5 printed along various edges. The picture of the object has letters along all its visible sides. Test takers imagine the paper folded to form the object and indicate which of the lettered edges on the object correspond to the numbered edges on the paper. There are 12 items in 2 parts. Test takers have 6 minutes to complete each part (i.e., a total of 12 minutes). Scores are computed as the number of edges correctly matched minus a correction fraction of the number of edges incorrectly matched.

This test is similar to the Paper Folding Test in requiring mental transformation of objects in three dimensions. This test is more of a power test. It could be easily computerized, and it is commercially available from the ETS. It is a fairly long, purely visual test.

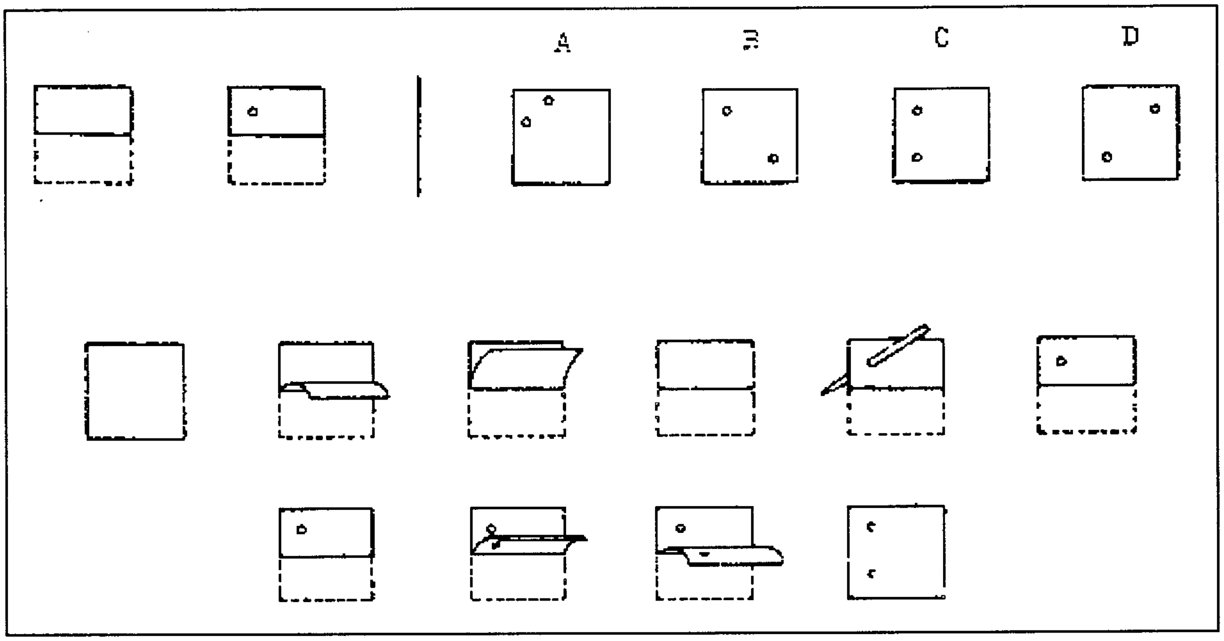


Figure 21. Example of items for the Paper Folding test.

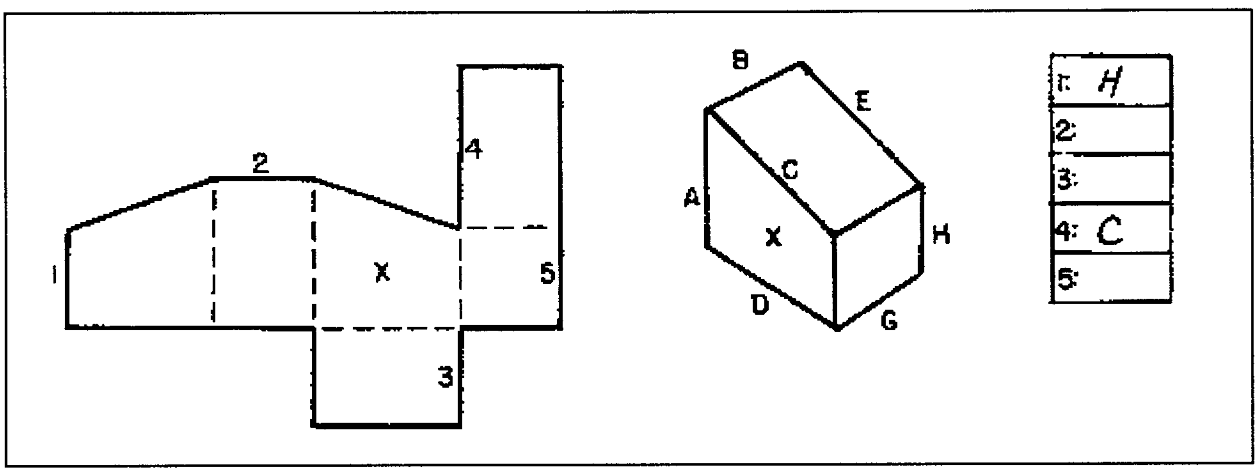


Figure 22. Example of items for the Surface Development test.

4.3.3.9 Matrix Rotation

This test is part of the Unified Tri-Service Cognitive Performance Battery (UTC-PAB) developed for the Naval Health Research Center (Englund et al., 1987; cited in Kennedy, et al., 1989). The computerized battery assesses general cognitive functioning. The Matrix Rotation Test measures spatial orientation ability, particularly mental rotation and spatial memory.

A series of five x five cell matrices are presented with five illuminated cells per matrix. Two successive matrices are never in the same orientation; a matrix is rotated either 90° to the left or right of the preceding matrix. Test takers compare successive displays and deduce if they are the same or different from the immediately preceding matrix. A matrix is the same as the preceding one if the same cells are illuminated but the matrix is rotated to the left or right. The item remains on the computer screen until the test taker responds. Scores are computed as the mean response time to identify matching matrices.

Tests in the UTC-PAB have an average reliability of .71 (Kennedy et al., 1989). A factor analysis generated a validity coefficient of .68 for the Matrix Rotation test with other tests of spatial ability. The test is already computerized and has been evaluated. It is primarily a power test with no time limit. It is a highly visual test. The length of the test was not reported, and the test may not be commercially available.

4.3.3.10 Spatial Antecedents Questionnaire (SAQ)

This test was developed by Olson and Eliot (1986) as a predictor of performance in spatial tests. Appendix B lists the items of the SAQ that correlate with performance on the Spatial Dimensionality Test, which also assesses visuo-spatial processing.

The SAQ is a multiple-choice measure with five subscales. The Activities Scale, containing 104 items, uses a 3-point scale to measure test takers' participation in different activities that require spatial abilities (e.g., various sports, visual games, drafting). The Academic Courses Scale, containing 21 items, uses a 2-point (yes/no) scale to measure participation in academic courses related to spatial ability (e.g., physics, geometry). The Self Assessment Scale, containing 15 items, uses a 5-point scale to measure how easily the test taker can perform tasks requiring spatial ability (e.g., rotating objects, understanding graphs, drawing things). The Environmental Mapping Scale contains 27 items in 2 parts. Part 1 contains 10 items designed to assess the identification of landmarks on the University of Maryland campus. Part 2 contains 12 items to measure distance estimates between landmarks, and 5 items to measure preferences for processing directional information (e.g., north-south, left-right of a landmark). The test is not timed, but test takers should require approximately 10 minutes.

The SAQ has been found to have correlations ranging from .23 to .44 with tests of spatial ability (Olson & Eliot, 1986). The Academic Courses and Self Assessment Scales accounted for most of the variability. The SAQ was found to have a reliability coefficient of .92, although the Environmental Mapping Scale has low reliability (.29). In general, the SAQ measures participation in spatial activities such as mental rotation, visualization, and spatial memory used in object recognition. The SAQ could be easily computerized. It would be a short test (approximately 5 minutes) if only the two most valid scales (Academic Courses and Self Assessment) were used. The test is a verbal questionnaire and only indirectly measures visuo-perceptual processes. The SAQ is not commercially available but could be obtained from Olson and Eliot.

4.3.3.11 Revised Individual Differences Questionnaire (IDQ)

This test was designed by Paivio (1971; cited in Paivio & Harshman, 1983) to measure imagery and verbal skills and preferences. A sample is depicted in Appendix C. The IDQ measures self-reported use of imagery and verbal strategies in a range of tasks.

There are 86 true-false statements. Items assert or deny preferences, abilities, and habits related to imaginal and verbal processing (e.g., "I use mental imagery or pictures to help me remember things"). Items are assigned to imagery and verbal scales. Scores for imaginal and verbal processing are calculated by giving a point for each consistent response in imagery and verbal scales.

The imagery items of the IDQ have a median reliability coefficient of .80, and the verbal items have a median reliability coefficient of .83 (Paivio & Harshman, 1983). Paivio and Harshman found that the imagery and verbal items correlated with different factors in a factor analysis. Twenty-three imagery items had factor loadings greater than .30 on a factor related to vividness of imagery and factor loadings less than .25 on any other factor. Thirty-one verbal items had factor loadings greater than .30 on a different factor related to verbal fluency and factor loadings less than .25 on any other factor. The IDQ assesses relative preference for an imaginal versus verbal cognitive style. It predicts how well a person is able to use imagery in perceptual tasks such as pattern integration and object recognition. The test is of moderate length, although it could be shortened by using only the imagery subscale. It is a verbal assessment but is designed to measure mental imagery not visual processing. It would be easy to computerize. The IDQ is not commercially available, but the test is published in an article by Paivio and Harshman. The Canadian Psychological Association holds the copyright to that article.

4.3.3.12 Picture Memory

This test was developed by DeFries, Plomin, Vandenberg, and Kuse (1981; cited in Kent & Plomin, 1987). It was designed as a test of immediate and long-term memory. However, it employs exclusively visual stimuli and taps memory processes related specifically to object identification and representation.

Items consist of line drawings of common objects. In the immediate test, test takers view 40 pictures for 45 seconds. Test takers then receive a set of 40 test items, 20 of which were target pictures and 20 of which are new distracters. Test takers indicate which test pictures were studied. The test takes approximately 5 minutes to administer. In the delayed test, test takers view 20 pictures for 1 minute. There is a retention interval of 15 minutes during which the test taker does other tasks. Then test takers receive a set of 20 test items, 10 of which were target pictures and 10 of which are new distracters. The delayed test requires approximately 30 minutes to administer.

The reliability for both tests is .65 (Kent & Plomin, 1987). Few data are available on the validity of the test. Kent and Plomin measured the effectiveness of a telephone version and found that the immediate and delayed tests correlated strongly with each other and did not correlate with verbal, spatial, or perceptual speed tests. The tests would be fairly easy to computerize, although the delayed test would have to allow the test taker to do other tasks during the retention interval. Both tests are short, although the delayed test has a 15-minute interval separating study and test phases. The test is purely visual. It is not commercially available and would have to be obtained through the authors.

4.3.3.13 Visual Recognition Test

This computerized test was developed by Juhel (1991) to study spatial imagery and memory abilities. It measures visualization and mental rotation abilities.

Figure 23 depicts examples of items for this test. Items consist of block shapes presented for 1 second. After a brief, 600 msec interval, a recognition set of four shapes is presented. Test takers must recognize the target shape in the set. There are 40 items in 2 parts. In the first part, the target is presented in the same orientation as during the study. In the second part, the target item is presented at a rotation of 90° clockwise.

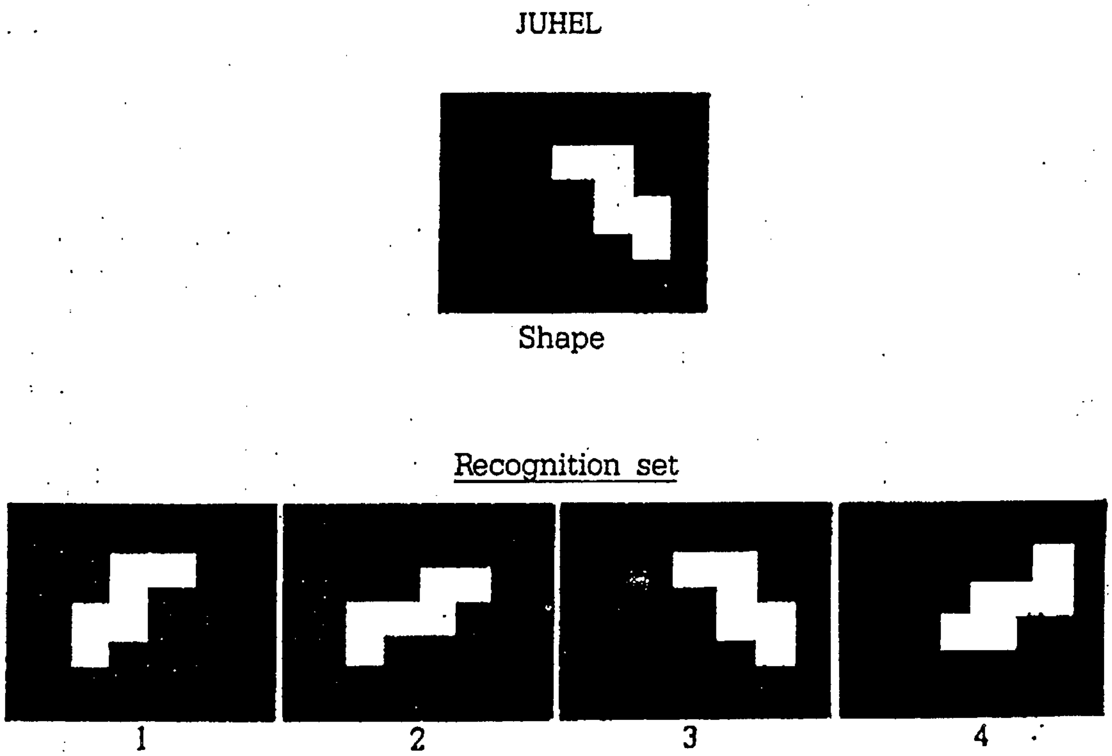


Figure 23. Example of items for the Visual Recognition test.

This test has a high correlation (.13 to .24) with tests of visualization and visual memory (Juhel, 1991). It correlates more weakly (.08) with the Cards Rotation Test which assesses spatial orientation. Overall, there are moderate indications of validity. No data on the reliability of the test are available. This test is computerized, short, and purely visual. It may not be commercially available.

4.3.4 Tests Related to Classification

4.3.4.1 Letter Sets

This test is produced by the ETS. It is designed to measure Induction, which is defined as "the kinds of reasoning abilities involved in forming and trying out hypotheses that will fit a set of data" (Ekstrom et al., 1976, p. 79). The skill encompasses concept formation and hypothesis testing, specifically forming groups of objects and verifying inclusion in a group.

Figure 24 depicts examples of items for this test. Items consist of five sets of letters with four letters in each set. Four sets are alike in some way (e.g., having three letters the same, being sequential letters of the alphabet). Test takers indicate which of the five sets of letters is unlike the others. The rule governing which sets are alike is different for each item. There are 30 items in 2 parts. Test takers have 7 minutes to complete each part. Scores are computed as the number of items correct minus a corrective fraction of the items incorrect.

A.	NOPQ	~~DEFL~~	ABCD	HIJK	UVWX
B.	NLIK	PLIK	QLIK	~~THIK~~	VLIK

Figure 24. Example of items for the Letter Sets test.

This test requires test takers to induce a categorical relationship among objects and judge category membership among instances. The test measures peoples' ability to find relationships among things. This test could be easily computerized and is commercially available from the ETS. It is an abstract classification task and is fairly long (i.e., 14 minutes).

4.3.4.2 Locations Test

This test is produced by the ETS and also measures Induction.

Figure 25 depicts examples of items for this test. Items consist of five rows of dashes separated into groups by blank spaces. In each of the first four rows of an item, one dash is replaced by an 'x.' In the fifth row, five of the dashes are replaced by numbers. Each item has a rule governing the placement of the 'x' in the first four rows. Test takers must induce the rule for each item and indicate the number in the fifth row that corresponds to the correct location where the 'x' should be. There are 28 items in 2 parts and test takers have 6 minutes to complete each part (i.e., 12 minutes total). Scores are computed as the number of items correct minus a corrective fraction of the items incorrect.

This test assesses classification, using visuo-spatial categories, which is relevant to aspects of the X-ray screening task. The items, however, are fairly abstract. This test could be easily computerized and is commercially available from the ETS.

```
Example A:

                 Row 1   -------------- --x -----
                 Row 2   ----- --x-- ------------
                 Row 3   ------------- --x ------
                 Row 4   ----- --x---- ----------
                 Row 5   -------1-- --2--3-4 5---


Example B:

                        -------- x----- --------
                        -- --------- -- x-- ----
                        x------- ---------------
                        ----------- x----- -----
                        1------2 3---- 4-- -5---
```

Figure 25. Example of items for the Locations test.

4.3.4.3 Figure Classification

This test is produced by the ETS and is designed to measure Induction.

Figure 26 depicts examples of items for this test. Items consist of two or three groups, each consisting of three figures and a set of test figures. The figures are simple visual patterns (e.g., geometric shapes, line patterns). All three figures in a group are alike in some respect (e.g., all are composed of straight lines or all are shaded). For each test figure, test takers indicate to which group the figure belongs. There are 28 items in 2 parts. Each item has eight test figures. Test takers have 8 minutes to complete each part, for a total of 16 minutes. Scores are computed as the number of test items correctly classified minus a fraction of the number incorrectly classified. It is primarily a power test.

This test assesses classification using purely visual categories which is relevant to aspects of the X-ray screening task. This test could be easily computerized and is commercially available from the ETS. It is a fairly long test.

4.3.4.4 Picture Number Test

This test is produced by the ETS and is designed to measure Associative Memory, which is defined as "the ability to recall one part of a previously learned but otherwise unrelated pair of items when the other part of the pair is presented" (Ekstrom et al., 1976, p. 93). Associative memory is only one component of successful classification, but it is a crucial component. An individual must have the ability to store a large number of objects in memory and associate objects to categories to quickly classify encountered objects.

Figure 27 depicts examples of items for this test. Test items consist of picture-number pairs. The pictures are line drawings of common objects. Test takers study a set of 21 items for 4 minutes. They are then presented with the pictures in a different order and write the

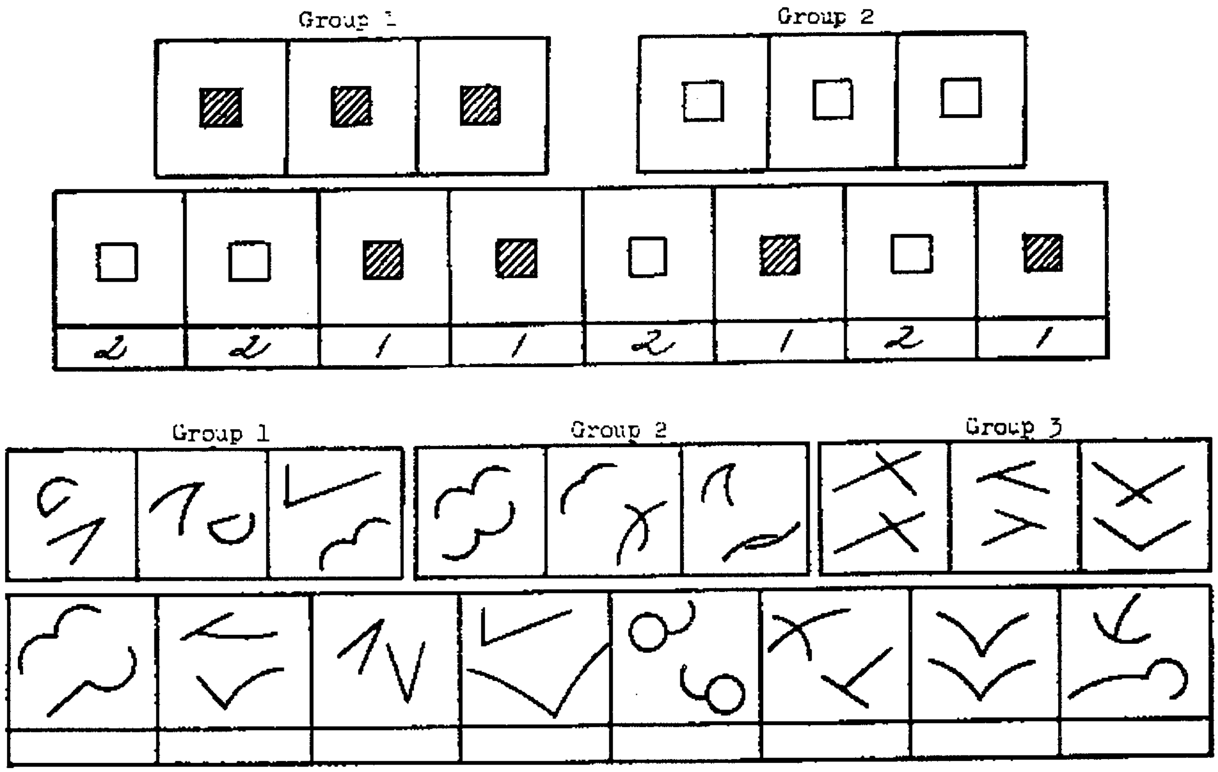


Figure 26. Example of items for the Figure Classification test.

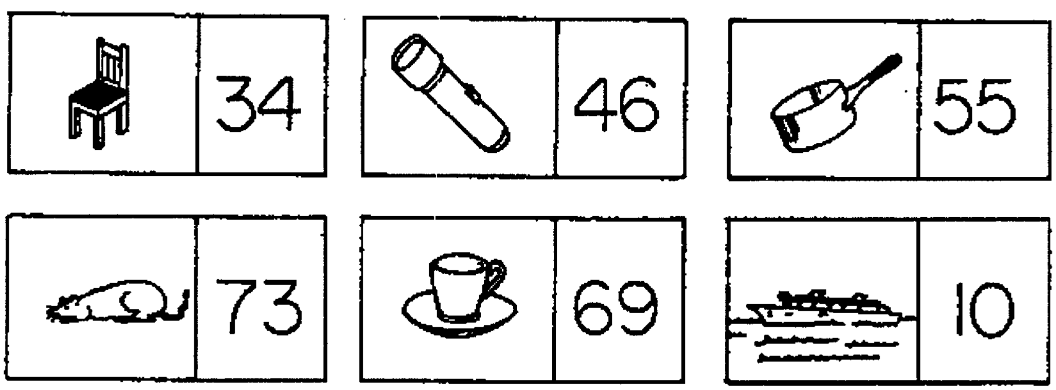


Figure 27. Example of items for the Picture Number test.

corresponding number next to the appropriate picture. There are 42 items in 2 parts. Test takers have 3 minutes for the recall phase of each part. It is a moderately long test at 15 minutes and is primarily a power test. Scores are computed as the number of associations correctly recalled.

This test assesses visual associative memory, although it probably overlaps with general associative abilities. The ability of X-ray screeners to learn associations of objects or patterns to categories is important. This test could be computerized, with test takers typing recall responses using a keyboard or numberpad. The test is commercially available from the ETS.

4.3.4.5 First and Last Names

This test is produced by the ETS and is designed to measure Associative Memory.

Figure 28 depicts examples of test items consisting of fictional names. Test takers study a set of 15 full names for 3 minutes. Test takers then receive the 15 last names in a different order and recall the first names paired with each. There are 30 items in 2 parts and test takers have 2 minutes to recall names in each part. It is a shorter test at 12 minutes total than the Picture Number Test and is a combination of power and speed tests. Scores are computed as the number of names correctly recalled.

Janet Gregory

Thomas Adams

Ronald Donaldson

Patricia Fletcher

Betty Bronson

Figure 28. Example of items for the First and Last Names test.

This test assesses associative memory but uses verbal materials so it is less relevant to X-ray screening than the Picture Number Test. This test could be computerized, but test takers would have to type responses using the keyboard and misspellings could be problematic. It is commercially available from the ETS.

4.3.4.6 Object Number Test

This test is produced and commercially distributed by the ETS. It is designed to measure Associative Memory.

Figure 29 depicts examples of test items consisting of pairs of words and numbers. Test takers study a set of 15 pairs for 3 minutes. They then receive the 15 words in a different order and recall the corresponding number for each. There are 30 items in 2 parts. Test takers have 2 minutes for the recall phase of each part. It is a moderately long test at 10 minutes total. Scores are computed by the number of items correctly recalled.

Object	Number
window	73
desk	41
carpet	19
door	84
glass	90

Figure 29. Example of items for the Object Number test.

This test assesses associative memory but uses verbal materials. This test could be computerized with test takers typing recall responses using a keyboard or numberpad.

4.3.4.7 Code Substitution Test

This test is part of UTC-PAB. It was adapted from the Wechsler Adult Intelligence Scale and is designed to measure associative learning ability and perceptual speed.

Test items consist of strings of letters and digits. A string of nine letters and nine digits are displayed on a computer screen. The digit string is shown directly below the letter string. Letters and digits are randomly paired for each item, and their order is randomly assigned in the coding string. A test letter is presented at the bottom of the screen below the coding strings. Test takers indicate which digit corresponds to the test letter. The number of items and length of the test were not available.

Tests in the UTC-PAB have an average reliability of .71 (Kennedy et al., 1989), which is not particularly high but is statistically reliable. Information on the validity of this test is limited. A factor analysis revealed correlations between .43 and .59 with tests of spatial processing, indicating that the test measures perceptual fluency in addition to categorization skill. The test is already computerized. It is an abstract test that uses verbal materials not directly related to classification in X-ray screening visual test. The test may not be commercially available.

4.3.5 Tests Related to Decision Making

4.3.5.1 State Trait Anxiety Inventory (STAI)

This test was developed by Speilberger, Gorsuch, Lushene, Vagg, & Jacobs (1983; cited in Mogg, Bradley, & Hallowell, 1994). It is designed to measure trait anxiety, which is the predispositional vulnerability to anxiety under stress. Anxious individuals selectively allocate processing resources to threat stimuli.

The STAI is a questionnaire that collects self-report of tendencies to experience anxiety, arousal, or fear in a variety of situations. The number of items and test length are not available.

This test would be easy to computerize, but it is a verbal measure. Anxiety and reaction to stress are not specific to visual stimuli, however. X-ray screeners need to be able to respond to stress and continue to allocate attentional resources effectively. The reliability and validity of the STAI are not available. This test is commercially available through the Consulting Psychologists Press.

4.3.5.2 Responsibility Scale

This test is part of the California Psychological Inventory (CPI) (cited in Weekes, 1993), produced by Consulting Psychologists Press. The CPI is a self-report questionnaire that measures 20 personality characteristics. The Responsibility Scale is designed to identify individuals who show a ready willingness to accept the consequences of their own behavior, dependability, trustworthiness, fiscal integrity, self-discipline, and sense of responsibility to self and others.

Items consist of statements regarding behavior in various situations. Test takers respond true or false to items to indicate relative risk taking. There are 36 items with no explicit time limit. Scoring is based on comparison of test takers' responses to established norms. The CPI as a whole includes 194 items from the Minnesota Multiphasic Personality Inventory that refer to symptoms of pathology. The items are unlikely to be offensive to test takers, but it is unclear that they assess variations in normal personality functioning.

People classified as responsible by the scale tend not to participate in risky behaviors (e.g., gambling, drug use). Gough (1987; cited in Weekes, 1993) reports a reliability coefficient of .71 and a concurrent validity of .38 for the scale with independent ratings of responsibility. Weekes reports a high predictive validity (.34) of the scale for search for new information in situations having risk consequences for the self. The Responsibility Scale should predict the tendency of an X-ray screener to pursue less risky decisions and to seek more information to make decisions. The test is short and could be easily computerized. It is a verbal inventory but assesses a general trait. The test is commercially available from Consulting Psychological Press.

4.3.5.3 Choice Dilemma Questionnaire

This test was developed by Kogan and Wallach (1964; cited in Weekes, 1993). It is designed to measure the amount of risk an individual would advise others to take in hypothetical situations. Thus, it measures responsibility and sense of willingness to take risks on behalf of others.

Items consist of hypothetical decision-making situations. For each item, there is a set of possible actions a person could take. Test takers indicate the action they would advise another person to take. The number of items and total length of the test are not available.

The split-half reliability of the test has been measured at .62, which is not very high. Blascovitch, Ginsburg, & Veach (1975; cited in Weekes, 1993) and Malamuth (1975; cited in

Weekes, 1993) report a high degree of predictive validity of the test in relation to risk-taking behavior in student counseling and gambling but do not report validity coefficients. The predictive validity of the test for actual risk-taking behavior was established in relation to student counseling advise and gambling (Blascovitch et al.; Malamuth; cited in Weekes). Generally, high scores on the test indicate greater willingness to take risks for others. It is a verbal test of general decision-making strategy. The test could be easily computerized, but it may not be commercially available.

4.3.5.4 Kirton Adaption-Innovation Inventory

This test was developed by Kirton (1976) to measure decision-making style (see Appendix D). It assesses the degree to which individuals seek 'paradigm-consistent' (i.e., rule-based, well learned procedures) versus 'paradigm-inconsistent' (i.e., innovative, self-generated procedures) approaches to solving problems. The test measures the ability to act consistently and follow established procedures in decision making. It is a short, verbal test of a general cognitive factor.

Each item asks the test taker to indicate how easy or hard it would be to behave consistently over a long period in a particular way. Some examples of items are "has original ideas," "copes with several new ideas at the same time," and "masters all details painstakingly." Test takers estimate the level of difficulty in being consistent with the behavior indicated in an item. There are 32 items, which use a 5-point scale to collect responses. There is no time limit, but typically takes 5 to 10 minutes. Each item is scored from 1 to 5, and scores can range from 32 (extreme adaptor; follow set procedures) to 160 (extreme innovator; produce own procedures).

The test is very reliable. Test-retest coefficients of .82 to .84 for inter-test intervals of 3 months to 3.5 years have been reported (Clapp, 1993; Gryskiewicz, Hills, Holt, & Hills, 1987; Kirton, 1978; cited in Clapp, 1993). The test is valid as well. Richards and Gaston (1995) report correlations of .71 to .81 between the Kirton Adaption-Innovation Inventory and self-report measures of decision-making style. Test scores also correlate with independent measures of factors such as conservatism, extroversion, tolerance of ambiguity, and cognitive flexibility (Kirton, 1976). Test scores also correlate with independent ratings of innovativeness (Keller & Holland, 1978; cited in Kirton & McCarthy, 1985) and self estimation of innovativeness (Kirton & McCarthy). The test has a high level of item transparency and is prone to motivational distortion such as faking and response bias. It is not clear exactly how this test relates to the X-ray screening task. Screeners must follow protocols but are also required to deal with novel situations. Presumably, someone scoring in the midrange of the scale, without an extreme decision-making style, would be preferred. The test does relate to two cognitive factors important to the screening process: a) the consistency of decision making and the ability to follow guidelines, and b) the acceptance of risk (high innovator scores would indicate greater acceptance of potential negative outcomes). The test is published in the United States by Mercatus, Inc. The test would be easy to computerize.

4.3.6 Tests Related to Attention and Vigilance

4.3.6.1 AGARD-STRES Reaction Time

This test is part of the AGARD Standardized Tests for Research with Environmental Stressors (AGARD-STES) battery (Reeves et al., 1991; cited in Draycott & Kline, 1996). It was developed by the Aerospace Medical Panel - Working Group 12 to measure ability to respond to stimuli under various levels of difficulty.

The test contains three 3-minute tasks, each a variation on the first. Stimuli in the first task are made up of single large digits that appear in the center of a computer screen one at a time. The interstimulus interval is 2 seconds following a correct response and 2.5 seconds following an incorrect response. Test takers press the left mouse button if the digit is '2' or '3' or the right mouse button if the digit is '4' or '5.' Items in the second task are the same digits, but they are degraded and harder to identify. Items in the third task are the same digits, but the interstimulus intervals are random, and the occurrence of items is difficult to predict. In all tasks, scores are computed as mean response time and accuracy of response.

The AGARD-STRES battery has a test-retest reliability coefficient of .76 to .89 (cited in Draycott & Kline, 1996). On the face, the Reaction Time test is primarily a measure of processing speed as well. It may measure more general cognitive processing. A factor analysis has revealed that the Reaction Time Test loads (factor loadings of .65 to .73) on a factor related to perceptual speed. Spatial processing, memory search, and mathematical processing tests of the AGARD-STRES battery also load on this factor (Draycott & Kline). The Reaction Time Test loads to a lesser extent (.19 to .36) on a factor related to visual tracking. Thus, the test is not just a measure of perceptual fluency. Although speculative, the test be able to assess concentration and control of processing. The test is computerized and visual, although it uses digits as stimuli. The test is of moderate length (total time of 9 minutes) and is available from the Naval Aerospace Medical Research Laboratory.

4.3.6.2 Boredom Proneness Scale (BPS)

This test was developed by Farmer and Sundberg (1986) to measure boredom as a trait. Trait boredom is the tendency to lose concentration or interest in tasks across work situations and past-times.

The BPS is a 28-item scale with items consisting of statements affirming or denying ability to maintain concentration on a task. Test takers respond yes or no to items. A point is assigned to each response indicative of the tendency toward boredom.

The BPS has an internal reliability coefficient of .79 and a test-retest coefficient of .83 (cited in Sawin & Scerbo, 1995). Sawin and Scerbo report that the test has predictive validity as a measure of interest and attention in the classroom and as a predictor of performance on a visual vigilance task, but do not give validity coefficient values. The BPS is short, requiring approximately 10 minutes to complete. It is a verbal test but measures a general cognitive trait. The BPS may not be commercially available.

4.3.6.3 Computer Administered Visual Attention Test (CA-VAT)

This test was developed by Arthur, Strong, and Williamson (1994) to be an easily administered computerized test of visual attention ability. It is based on the Visual Selective Attention Test (VSAT), which was designed to predict performance on perceptual/cognitive tasks. In particular, these tests measure the ability to direct and maintain visual attention.

Stimuli are pairs of numbers and letters that appear on a computer monitor. A given pair of characters consists of either two letters, two numbers, or a number and a letter. Stimuli are presented in 24 series of 24 stimuli. The test taker's task is to respond to stimulus pairs via the computer keyboard. Cue words, presented prior to a series, indicate the appropriate responses for a series. The cues are arbitrary words that are associated with computer keys used to indicate the kind of stimulus pair. The word 'coffee' is used to signal that the left-arrow key is to be used to respond to all odd numbers at the left side of the display and the right-arrow key to all even numbers at the right side of the display. The word 'apple' is used to signal that the left-arrow key is to be used to respond to all even numbers at the left side of the display and the right-arrow key to all odd numbers at the right side of the display. Test takers are never to respond to letters, which are distracters. Changing the cues requires test takers to change their responses to stimuli. Scores are computed as the number of errors made. The test takes approximately 25 minutes.

The CA-VAT has test-retest coefficient of .83 (Strong, 1992; cited in Arthur et al., 1994). The auditory version of this test and the VSAT have been found to predict performance in flying, monitoring, and driving. Three versions of the CA-VAT are significantly correlated (.38, .28, and .26) with accident involvement for a sample of volunteer drivers (Arthur et al.). The task requires control of decision processes about visual stimuli. It seems suitable for domains such as X-ray screening involving perception and identification of multiple stimuli and responding to the most important aspects of the information. The test is computerized and assesses visual attention, although it uses alphanumeric characters as stimuli. The test is long, although a shorter version exists (half as long). The CA-VAT may not be commercially available.

4.3.6.4 Stroop Color Word Test

This test is based on Stroop's (1935) demonstration of interference among simultaneous cognitive processes. The test is designed as a measure of attentional deficit (see Batchelor, Harvey, & Bryant, 1995). It measures focused attention, the ability to direct processing to important stimuli while excluding irrelevant information.

There are three parts to the test. In the word subtask, test takers read a series of color names (blue, red, yellow, etc.; the colors are repeated) printed in black ink. In the color subtask, test takers name the color of triangles. In the color-word subtask, test takers name colors of color names printed in conflicting color (e.g., the word 'red' printed in blue ink). There are 100 items per subtest. Scores are computed as the difference in time to complete the color word and color subtasks. Completion time in the color names subtask can be used to adjust scores for reading time.

Reliability coefficients for the group form and individual subtasks range from .69 to .89 (Hynd, 1985). Data on the validity of the test are not available. The test possesses some face validity as a measure of attention in a task such as X-ray screening. The test requires maintenance of focused attention to overcome interference. The Stroop test has been noted as most useful for diagnosis of brain dysfunction, and some suggest it has limited applicability for normal populations (Evans, 1985). The test could be easily computerized and is of moderate length (approximately 15 minutes to administer). The test is largely visual but also involves reading. It may not be commercially available.

4.3.6.5 Beck Depression Inventory

This test was developed by Beck (1967; cited in Ingram, Bernet, & McLaughlin, 1994) to measure depressive symptomatology. Behaviors associated with depressive states include negative self-evaluation and loss of interest in work and past-times. Importantly, dysfunctional information processing and impaired cognitive functioning are also associated with depressive symptoms (see Ingram et al.).

Items consist of statements affirming depressive symptoms and behaviors (e.g., "I frequently find it difficult to concentrate"). Test takers rate the applicability of items to themselves using a four-point (0-3) scale of severity. Each item has four sentences, ranging from no complaint to severe complaint (e.g., "0 - I do not feel sad" to "3 - I am so sad or unhappy that I can't stand it"). The first 13 items cover cognitive-affective forms of depression, on topics such as pessimism, guilt, and indecisiveness. Eight items cover somatic-performance aspects of depression, on such topics as work and sleep difficulties and loss of interest in sex. There are 21 items and scores are computed as the total number of points assigned to the 21 items (0 to 63 total). The test is self-paced but typically requires 5 to 10 minutes.

The reliability of the test is high, with an internal consistency of .81 to .88 (Beck, 1967; Beck, Steer, & Garbin, 1988; cited in Ingram et al., 1994). Test-retest correlation coefficients range from .60 to .83. The test has a correlation coefficient of .72 with clinical ratings of depression for psychiatric samples and a coefficient of .60 for individuals not undergoing psychiatric treatment. The test is aimed at and used with clinical populations. Concurrent validity coefficients range from .60 to .76. Even though it is a valid predictor of impaired information processing, the Beck Depression Inventory may not be sensitive to differences in a non-clinical population of potential X-ray screeners. It is a verbal assessment of general cognitive functioning. The test is easy to computerize and commercially available from The Psychological Corporation.

4.3.6.6 X-Ray Image Monitoring Task

Given the relative paucity of tests to assess visual attention and vigilance ability, it may be useful to develop a test specifically for that purpose. Such a test should be designed to measure selective visual attention ability in a situation close to the X-ray screening environment.

Items could consist of X-ray or X-ray-like images. Each item could contain clear target objects and a set of visually similar distracters. Because this is a selection test, the images do not have to be exactly like the X-ray images seen on the job and the targets and distracters do not have to appear like normal objects in carry-on luggage. Test takers should monitor a sequence of images for the presence of a predetermined target. The target could appear at a set frequency (e.g., 50% of trials) for a block of items. Test takers would signal the presence of a target as quickly and accurately as possible. To keep attention requirements high, the target item should be changed from one block to the next. A former distracter object could become the target for the next block, and the former target could become a distracter. The length of the test could be adjusted to fit constraints of reliability and practicality. There might be, for example, a total of 128 items in 4 blocks of 32 items each (16 targets and 16 non-targets). Scores could be computed by reaction time and/or error rate.

This test should be computerized and purely visual. It should be of moderate length, no more than 12 minutes. The validity and reliability of the test will have to be empirically established. This test is primarily a power test.

5. Recommendations

Table 4 presents a list of 14 tests that are recommended as potentially valid and useful X-ray screener selection tests. The cognitive model of X-ray screening allows the construction of a battery of selection tests that will assess abilities relevant to all stages of the screening process. The set of recommended tests covers all the major cognitive processes identified in the model. Thus, the battery should be a valid predictor of X-ray screener proficiency.

The recommended tests were chosen to meet the criteria for a useful selection test. These tests are all valid and reliable predictors of particular cognitive abilities. Further, they are practical for use in selection of X-ray screeners. The tests can be computerized and would not require any specialized equipment. There is a mix of power and speed tests throughout the set. The recommended tests are generally short (i.e., less than 10 minutes), although a few tests are of moderate length (i.e., 10 to 15 minutes). Overall, the 14 recommended tests sum to greater than 60 minutes, but some tests will likely be removed if they prove insufficiently valid or reliable as predictors of X-ray screening performance or difficult to implement. The recommended tests emphasize visual processing and typically use pictorial stimuli. Some cognitive processes in X-ray screening are general and difficult to assess in the visual domain, so several rating scales and non-visual tests are included. All the tests should be available with minimal difficulty.

The following is a brief synopsis of each of the 14 recommended tests for X-ray screener selection. The Identical Pictures Test seems very useful because it measures speed and accuracy in responding to basic visual features. This test should measure ability in visual search and feature extraction. Other tests that measure the same processes (e.g., Finding As and Number Comparison) use alphanumeric characters and are less similar to the screening environment.

The Choosing a Path Test is also valuable. It measures visual scanning and visual attention across space. This test is fairly long at 14 minutes, reducing its practicality. The Map Planning Test may be more convenient, but it is also a more complex test.

Table 4. Recommended Selection Tests

Recommended Test	**Relevant Cognitive Process**	**Type Of Test**	**Length***
Identical Pictures	Image Generation	Speed/Power	3 min.
Choosing a Path	Image Generation	Power	14 min
Simultaneous Pattern Comparison	Image Generation	Speed	1.5 min
Hidden Patterns	Pattern Integration	Speed	6 min
Snowy Pictures	Pattern Integration	Power	8 min
VVIQ	Pattern Integration	Questionnaire	approx. 2 min
Three Dimensional Space	Pattern Integration	Power	Unknown
Cube Comparison	Object Recognition	Speed/Power	6 min
Paper Folding	Object Recognition	Speed/Power	6 min
Matrix Rotation	Object Recognition	Power	Self-paced
Picture Number Test	Classification	Power	15 min
Locations Test	Classification	Power	12 min
Responsibility Scale	Decision Making	Questionnaire	Self-paced
Boredom Proneness	Attention & Vigilance	Questionnaire	approx. 10 min

* Not including time to give instructions

The Simultaneous Pattern Comparison Test is short and visual. It tests the ability to locate and respond to distinctive visual features. The features and patterns are not pre-specified in this test, so it complements the other recommended tests. The Reaction Time test is also potentially useful but tests primarily the speed of responding. The Tool Matching Test is also a potentially valid test, but it overlaps to some degree with tests of pattern integration.

The Hidden Patterns Test assesses abilities to integrate features into a pattern and to identify patterns in a complex field. The Hidden Figures Test measures the same capacities but may be a preferred test because it uses stimuli closer to natural objects. Practically, the Hidden Figures Test may not be feasible because, at 24 minutes, it is substantially longer than the 6-minute Hidden Patterns Test. Likewise, the Copying Test is impractical because it would require special computer equipment to collect responses.

The Snowy Pictures Test measures the ability to mentally construct an unknown pattern from ambiguous visual elements. It has high face validity for X-ray screening, where the screener must identify shapes and objects from a degraded picture. There may be some problem collecting responses in a computerized version. A procedure for dealing with potential spelling errors will have to be developed. The Gestalt Completion Test may also be a good test of pattern integration should the Snowy Pictures Test prove impractical. The Concealed Words Test employs the same processes but may have lower validity because it uses verbal stimuli and may not fully engage visual processing.

The VVIQ is a widely used measure of visual imagery ability. To the extent that imagery is used in perceptual processing, this scale will indicate screening performance. The VVIQ is very easy to administer and very short (2 minutes). Some subscales of the SAQ may also be useful measures of imagery ability.

The Three Dimensional Space Test measures the perceptual identification of 3D forms. It is shorter by at least 5 minutes and more convenient than the Paper Folding and Surface Development Tests. Those tests may be reasonable replacements if the Three Dimensional Space Test is not commercially available.

The Cube Comparison Test provides a validated assessment of 3D mental rotation ability, which is crucial to object recognition. The Card Rotation Test could also be useful, although it is limited to rotation in 2D.

The Paper Folding Test is recommended because it measures transformation ability related to inducing the 3D structure of objects from 2D images. This ability is very pertinent to X-ray screening, which depends on interpretation of flat images of objects. The Surface Development Test is longer by 6 minutes and somewhat more complex, although it does employ more natural 3D forms.

The Matrix Rotation Test assesses visual memory and rotational processes. X-ray screeners must possess good visual memory to learn to identify objects from X-ray images. The Picture Memory Test is a somewhat longer test of visual memory, taking 5 minutes to administer, and it would be a more difficult test to administer.

The Picture Number Test assesses associative memory, which is needed for identification and classification of X-ray images. This ability is closely related to visual memory. As a visual test, it has more face validity than the First and Last Name and Object Number Tests.

The Locations Test assesses the ability to induce category structure of novel stimuli. Items are spatial patterns, which are the closest stimuli to visual patterns available in tests of classification.

The Responsibility Scale measures a person's willingness to engage in risky behavior. The test should predict the effort screeners put into making accurate decisions that emphasize public safety. The Choice Dilemma Questionnaire, if available, could also be a useful test. The STAI emphasizes clinical levels of anxiety and would be less effective in measuring decision-making style in screeners. The Kirton Adaption-Innovation Test may be a reliable measure for screeners, but the ability to follow set procedures is not the most central feature of the screening job.

The Boredom Proneness Inventory should provide indication of how well a person will maintain attention in the screening task. Vigilance has traditionally been considered a very important aspect of the screener's job. The CA-VAT is a long test (i.e., 25 minutes) and requires complex instructions. It may be worthwhile to create an attentional monitoring task specifically for the X-ray screening environment. Such a test should allow behavioral measurement of attentional ability.

References

Advisory Group for Aerospace Research and Development, Aerospace Medical Panel Working Group 12. (1989). *Human performance assessment methods* (AGARDograph No. 308). Neuilly-sur-Seine, France: Author.

Anastasi, A. (1988). *Psychological testing* (6th Ed.). New York, NY: Macmillan.

Annett, M. (1992). Spatial ability in subgroups of left- and right-handers. *British Journal of Psychology, 83,* 493-515.

Arthur, W., Strong, M. H., & Williamson, J. (1994). Validation of a visual attention test as a predictor of driving accident involvement. *Journal of Occupational and Organizational Psychology, 67,* 173-182.

Baddeley, A. (1990). *Human memory: Theory and practice*. Boston, MA: Allyn and Bacon.

Batchelor, J., Harvey, A. G., & Bryant, R. A. (1995). Stroop Colour Word Test as a measure of attentional deficit following mild head injury. *The Clinical Neuropsychologist, 9,* 180-186.

Beck, A. T. (1967). *Cognitive therapy and the emotional disorders*. New York, NY: International University Press.

Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77-100.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115-147.

Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real world scenes. *Journal of Experimental Psychology, 97,* 22-27.

Blascovich, J., Ginsburg, G. P., & Veach, T. L. (1975). A pluralistic explanation of choice shifts on the risk dimension. *Journal of Personality and Social Psychology, 31,* 422-429.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

Brooks, L. R. (1967). The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology, 19,* 289-299.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Bryant, D. J., Reed, C., & Lanca, M. (1997). *Pattern recognition in vision and touch: Reliance on external and internal geometries of objects*. Submitted manuscript.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Clapp, R. G. (1993). Stability of cognitive style in adults and some implications: A longitudinal study of the Kirton Adaption-Innovation Inventory. *Psychological Reports, 73*, 1235-1245.

Corballis, M. C., & Nagourney, B. A. (1978). Latency to categorize disoriented alphanumeric characters as letters or digits. *Canadian Journal of Psychology, 32*, 186-188.

Corballis, M. C., Zbrodoff, N. J., Shetzer, L. I., & Butler, P. B. (1978). Decisions about identity and orientation of rotated letters and digits. *Memory & Cognition, 6*, 98-107.

Daniel, M. (1978). *Wiggly Block worksample 3, 4, 5, 6, 7 manual.* Boston, MA: Johnson O'Connor Research Foundation.

DeFries, J. C., Plomin, R., Vandenberg, S. G., & Kuse, A. R. (1981). Parent-offspring resemblance for cognitive abilities in the Colorado Adoption Project: Biological, adoptive, and control parents and one-year-old children. *Intelligence, 5*, 245-277.

de Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton.

de Groot, A. D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving: Research, method and theory.* New York: Wiley.

Draycott, S. G., & Kline, P. (1996). Validation of the AGARD STRES battery of performance tests. *Human Factors, 38*, 347-361.

EG&G Astrophysics Research Corporation (1994). *E-Scan X-ray system instructor's guide: Small parcel inspection.*

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Eley, M. G. (1982). Identifying rotated letter-like symbols. *Memory & Cognition, 10*, 25-32.

Englund, C. E., Reeves, D. L., Shingledecker, C. A., Thorne, D. R., Wilson, K. P., & Hegge, F. W. (1987). *Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB): I. Design and specification of the battery* (Report No. 87-10). San Diego, CA: Naval Health Research Center.

Evans, J.R. (1985). Review of Stroop Color and Word Test. In J. V. Mitchell (Ed.). *The Ninth Mental Measurements Yearbook* (pp 1484-1485). Lincoln, NB: University of Nebraska Press.

Farmer, R., & Sundberg, N. D. (1986). Boredom proneness - The development and correlates of a new scale. *Journal of Personality Assessment, 50*, 4-17.

Federal Aviation Administration. (1991, February). *Semiannual Report to Congress on the Effectiveness of the Civil Aviation Security Program* (Tech. Report DOT/FAA/ACS-89-1[31]). Washington, DC: NTIS.

Fobes, J. L., Cormier, S. M., McAnulty, D. M., & Klock, B. A. (1996). *Operational assessment for screener proficiency evaluation and reporting system (SPEARS) threat image projection* (DOT/FAA/AR-96/27). Atlantic City International Airport, NJ: FAA Technical Center.

Fobes, J. L., Lofaro, R. J., Malone, R. L., Fischer, D. S., & Berkowitz, J. (1994). *Test and Evaluation Plan (TEP) for Improvised Explosive Device Screening Systems* (IEDSS) (DOT/FAA/CT-94/87). Atlantic City International Airport, NJ: FAA Technical Center.

Fobes, J. L., & McAnulty, D. M. (1995). *Initial development of selection instruments for evaluating applicants as potential airline passenger screeners for conventional X-ray technology* (DOT/FAA/CT-95/41). Atlantic City International Airport, NJ: FAA Technical Center.

General Accounting Office. (1987, April). *FAA preboard passenger screening test results* (GAO/RCED-81-125FS). Washington, DC.

Galotti, K. M. (1994). *Cognitive psychology in and out of the laboratory*. Pacific Grove, CA: Brooks/Cole.

Gough, H. G. (1987). *California Psychological Inventory, administrator's guide*. Palo Alto, CA: Consulting Psychologists Press.

Gryskiewicz, S. S., Hills, D. W., Holt, K., & Hills, K. (1987). *Understanding managerial creativity: The Kirton Adaption-Innovation Inventory and other assessment measures* (Technical Report). Center for Creative Leadership, Greensboro, NC.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology, 166*, 106-154.

Hynd, G. W. (1985). Review of Stroop Color and Word Test. In J. V. Mitchell (Ed.). *The Ninth Mental Measurements Yearbook* (pp 1485-1486). Lincoln, NB: University of Nebraska Press.

Improvised munitions black book (1981). El Dorado, AZ: Desert Publications.

Ingram, R. E., Bernet, C. Z., & McLaughlin, S. C. (1994). Attentional allocation processes in individuals at risk for depression. *Cognitive Therapy and Research, 18*, 317-332.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13*, 289-303.

Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology, 42*, 461-478.

Jolicoeur, P., & Milliken, B. (1989). Identification of disoriented objects: Effects of context of prior presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 200-210.

Juhel, J. (1991). Spatial abilities and individual differences in visual information processing. *Intelligence, 15*, 117-137.

Kaempf, G., Klinger, D, Wolf, S., & Lofaro, R. J. (1994). *Development of decision-centered interventions for airport security checkpoints* (DOT/FAA/CT-94/27). Fairborn, OH: Klein Associates, Inc.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Keesling, J. W. (1985). Review of USES General Aptitude Test Battery. In J. V. Mitchell (Ed.). *The Ninth Mental Measurements Yearbook* (pp 1645-11647). Lincoln, NB: University of Nebraska Press.

Keller, R. T., & Holland, W. E. (1978). A cross-validation study of the Kirton Adaption-Innovation Inventory in three research and development organizations. *Applied Psychological Measurement, 2*, 563-570.

Kennedy, R. S., Baltzley, D. R., & Turnage, J. J. (1989). Factor analysis and predictive validity of microcomputer-based tests. *Perceptual and Motor Skills, 69*, 1059-1074.

Kennedy, R. S., Wilkes, R. L., Dunlap, W. P., & Kuntz, L. A. (1987). *Microbased repeated-measures performance testing and general intelligence*. Paper presented at the 29th Annual Conference of the Military Testing Association, Ottawa, Ontario, Canada.

Kent, J., & Plomin, R. (1987). Testing specific cognitive abilities by telephone and mail. *Intelligence, 11*, 391-400.

Kirton, M. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology, 61*, 622-629.

Kirton, M. J. (1978). Have adaptors and innovators equal levels of creativity? *Psychological Reports, 42*, 695-698.

Kirton, M., & McCarthy, R. M. (1985). Personal and group estimates of the Kirton Inventory scores. *Psychological Reports, 57*, 1067-1070.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York, NY: Harcourt Brace.

Kogan, N., & Wallach, M. A. (1964). *Risk taking: A study in cognition and personality*. New York, NY: Holt, Rinehart & Winston.

LaBerge, D. (1973). Attention and the measurement of perceptual learning. *Memory & Cognition, 1*, 268-276.

Lofaro, R. J., Gibb, G. D., Garland, D. J., Guide, P. C., Baker, J. S., & St. Laurent, M. (1994a). *A protocol for selecting airline passenger baggage screeners* (DOT/FAA/CT-94/107). Atlantic City International Airport, NJ: FAA Technical Center.

Lofaro, R. J., Gibb, G. D., Garland, D. J., Guide, P. C., Baker, J. S., & St. Laurent, M. (1994b). *Review of the literature related to screening airline passenger baggage* (DOT/FAA/CT-94/108). Atlantic City International Airport, NJ: FAA Technical Center.

Maki, R. H. (1986). Naming and locating the tops of rotated pictures. *Canadian Journal of Psychology, 40*, 368-387.

Malamuth, N. M. (1975). *A systematic analysis of the relationship between group shifts and characteristics of the Choice Dilemma Questionnaire*. Unpublished doctoral dissertation, University of California, Los Angeles.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64*, 17-24.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society, London, B, 200*, 269-294.

McDonald, R. A., & Eliot, J. (1987). The Wiggly Block Worksample and its relation to age, sex, and general intelligence. *Perceptual and Motor Skills, 65*, 675-678.

McMullen, P. A., & Jolicoeur, P. (1992). The reference frame and effects of orientation on finding the tops of rotated objects. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 807-820.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35,* 113-138.

Mogg, K., Bradley, B. P., & Hallowell, N. (1994). Attentional bias to threat: Roles of trait anxiety, stressful events, and awareness. *Quarterly Journal of Experimental Psychology, 47A,* 841-864.

Naglieri, J. A. (1985a). *Matrix analogies test - short form.* Columbus, OH: Charles E. Merrill.

Naglieri, J. A. (1985b). *Matrix analogies test - expanded form.* Columbus, OH: Charles E. Merrill.

Naglieri, J. A., & Insko, W. R. (1986). Construct validity of the Matrix Analogies Test - expanded form. *Journal of Psychoeducational Assessment, 4,* 243-255.

O'Connor, H. (1928). *Born that way.* Baltimore, MD: Williams & Wilkins.

Olson, D. M., & Eliot, J. (1986). Relationships between experiences, processing style, and sex-related differences in performance on spatial tests. *Perceptual and Motor Skills, 62,* 447-460.

Paivio, A. (1971). *Imagery and verbal processes.* New York, NY: Rinehart & Winston.

Paivio, A., & Harshman, R. (1983). Factor analysis of a questionnaire on imagery and verbal habits and skills. *Canadian Journal of Psychology, 37,* 461-483.

Posner, M. I., & Snyder, C. R. R. (1974). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium.* Hillsdale, NJ: Erlbaum.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160-174.

Poulton, E. C. (1977). Arousing stresses increases vigilance. In R. R. Mackie (Ed.), *Vigilance: Theory, operational performance, and physiological correlates* (pp. 423-459). New York, NY: Plenum.

Reeves, D. L., Winter, K. P., LaCour, S. J., Raynsford, K. M., Vogel, K., & Grissett, J. D. (1991). *The UTC-PAB/AGARD STRES battery: User's manual and system documentation* (Tech Report NAMRL SR90-1). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Reicher, G. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81,* 275-280.

Richards, T., & Gaston, K. (1995). A re-examination of personal and group estimates of the Kirton Adaption-Innovation scores. *Psychological Reports, 77*, 491-498.

Richardson, A. (1995). Guidelines for research with the VVIQ. *Journal of Mental Imagery, 19,* 171-177.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning.* Cambridge University Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192-233.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

Sawin, D. A., & Scerbo, M. W. (1995). Effects of instruction type and boredom proneness in vigilance: Implications for boredom and workload. *Human Factors, 37,* 752-765.

Shepard, R. N., & Cooper, L. A. (Eds.) (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Sternberg, R. J. (1996). *Cognitive psychology.* Orlando, FL: Harcourt Brace.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

Strong, M. H. (1992). *An assessment of the criterion-related validity and improvements in the utility of the Computer Administered Visual Selective Attention Test.* Unpublished Master's Thesis. Texas A&M University.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2,* 55-82.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21,* 233-282.

Taylor, E. M. (1959). *The appraisal of children with cerebral deficits.* Cambridge, MA: Harvard University Press.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97-136.

Treisman, A. M., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 459-478.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 4,* 207-232.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32,* 193-254.

U.S. Congress, House Committee on Government Operations. Government Activities and Transportation Subcommittee (1989). *The bombing of Pan Am Flight 103: A critical look at American aviation security.* Hearings before the Government Activities and Transportation Subcommittee on Government Operations, House of Representatives. Washington, DC: U.S. G.P.O., iv, 489, Doc. Numbers: GPO Item No 1016-B(MF).

Van de Vijer, F. J. R., & Harsveld, M. (1994). The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test Battery. *Journal of Applied Psychology, 79,* 852-859.

Van Hutton, V. (1990). Test review: The California Psychological Inventory. *Journal of Counseling and Development, 69,* 75-77.

Wagemans, J., Van Gool, L., & Lamote, C. (1996). The visual system's measurement of invariants need not itself be invariant. *Psychological Science, 7,* 232-236.

Weekes, B. S. (1993). Criterion-related validity of the responsibility scale of the California Psychological Inventory. *Psychological Reports, 73,* 315-320.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1,* 59-85.

Appendix A
Vividness of Visual Imagery Questionnaire (VVIQ)

IMAGERY QUESTIONNAIRE

The following questions will ask you to imagine certain people or objects. With your **eyes closed**, please try your best to perform the task and then rate your experience (1-5), using the following rating scale.

RATING SCALE:

1. Perfectly clear and as vivid as normal vision
2. Clear and reasonable vivid
3. Moderately clear and vivid
4. Vague and dim
5. No image at all; I only know that I am thinking of the object

	Rating
For items 1-4, think of some relative or friend whom you frequently see and consider carefully the picture that comes before your mind's eye	
1. The exact contour of face, head, shoulders, and body.	______
2. Characteristic poses of the head, attitudes of the body, etc.	______
3. The precise carriage, length of step, etc., in walking.	______
4. The different colours worn in some familiar clothes.	______
For items 5-8 visualize a rising sun. Consider carefully the picture that comes before your mind's eye.	
5. The sun is rising above the horizon into a hazy sky.	______
6. The sky clears and surrounds the sun with blueness.	______
7. Clouds. A storm blows up, with flashes of lightening.	______

8. A rainbow appears. ______

For items 9-12, think of the front of a shop which you often go to.

Consider the picture that comes before your mind's eye.

9. The overall appearance of the shop from the opposite side of the road. ______

10. A window display including colours, shapes, and details of the individual items for sale. ______

11. You are near the entrance. The colour, shape, and details of the door. ______

12. You enter the shop and go to the counter. The counter assistant serves. Money exchanges hands. ______

Finally, you think of a country scene which involves trees, mountains and a lake.

Consider the picture that comes before your mind's eye.

13. The contours of the landscape. ______

14. The colour and shape of the trees. ______

15. The colour and shape of the lake. ______

16. A strong wind blows on the trees and on the lake causing waves. ______

Appendix B

Spatial Antecedents Questionnaire Items that Significantly Correlate With Performance on the Spatial Dimensionality Test

Activities

Knitting
Tailoring
Mechanical Drawing
Play Chess
Tackle Football
Ballet
Choreograph Dance
High Jump
Running Hurdles
Car Repair
Solve Mathematical Riddles
Basketball
Canoeing
Sketching House Plans
Using Compass
Horseshoes
Making Jewelry
Photography
Carpentry
Advanced Racketball
Sketch Car Designs
Electrical Circuitry
Knot Tying
Frisbee
Paper Folding
Calligraphy
Building Models
Soccer
Beginning Tennis
Skiing
Doing Newspaper Layout
Play Pinball Games
Skiing
Baseball
Gymnastics
Drawing

Academic Courses

Calculus
Applied Design
Engineering Science
Astronomy
Physics
Architecture
Trigonometry
Drafting
Computer Programming
Biochemistry
Studio Arts
Electrical Engineering
Geography

Self-assessments

Arrange Objects
Draft/Draw Things
Rotate a Cube
Unfamiliar Building
Compass Directions
Understand Math/Science
Visualize Map of Pennsylvania
Work with Machines
Participate in Sports
Understand Graphs/Charts
Set-up Displays
Play Visual Games
Work on Computer Terminal

Appendix C
Revised Individual Differences Questionnaire

1. I have no difficulty in expressing myself verbally.
2. Listening to someone recount his experiences does not usually arouse mental pictures of the incidents being described
3. When reading fiction I usually form a mental picture of a scene or room that has been described.
4. Essay writing is difficult for me.
5. By using mental pictures of the elements of a problem, I am often able to arrive at a solution.
6. I enjoy being able to rephrase my thoughts in many ways for variety's sake when both writing and speaking.
7. I enjoy visual arts, such as paintings, more than reading.
8. I tell jokes and stories poorer than most people.
9. I enjoy doing work that requires the use of words.
10. My daydreams are something so vivid I feel as though I actually experienced the scene
11. I often use mental pictures to solve problems.
12. I enjoy reading an interesting story even if it is not particularly well written.
13. I find it difficult to find enough synonyms or alternate forms of a word when writing.
14. I have difficulty in expressing myself verbally.
15. My knowledge and use of grammar needs much improvement.
16. I would rather work with ideas than words.
17. I memorize material largely by the use of verbal repetition.
18. I enjoy learning new works and incorporating them into my vocabulary.
19. I do not have a vivid imagination.
20. I can easily picture moving objects in my mind.
21. Most of the time my thinking is verbal, as though talking to myself.
22. If given the choice, I would rather listen to a good speaker than visit an art gallery.
23. I find that I am more critical of writing style than content when reading literature.
24. I can form mental pictures to almost any word.
25. I have only vague visual impressions of scenes I have experienced.
26. My vocabulary is not as large as I would like.
27. When doing mental arithmetic, such as addition, I think in abstract terms rather than actually picturing the numbers
28. I can easily think of synonyms for words.
29. I think that most people think in terms of mental pictures whether they are completely aware of it or not.
30. I am able to express my thoughts clearly.
31. I remember things I have done myself, much better than things I have read.
32. My powers of imagination are higher are higher than average.
33. I consider myself a fast reader.
34. I have a large vocabulary.

35 I find it easy to visualize the faces of people I know.

36. My marks have been hampered by inefficient reading.
37. It bothers me when I see a word used improperly.
38. I don't believe that anyone can think in terms of mental pictures.
39. I can easily form a mental picture of Prime Minister Trudeau.

40. I am fluent at writing essays and reports.
41. I would rather have a verbal description of an object or person, than a picture.
42. I can close my eyes and easily picture a scene I have experienced.
43. I have a photographic memory.
44. I feel a picture of a friend's face when I close my eyes.
45. I cannot generate a picture of a friend's face when I close my eyes.
46. When someone describes something that happens to him, I sometimes find myself vividly imagining the events that happened.
47. I can add numbers by imagining them to be written on a blackboard.
48. I have found it easy in the past to learn a second language.
49. When I hear or read a word, a stream of other words often comes to mind.
50. I seldom dream.
51. I read rather slowly.
52. I am usually able to say what I mean in my first draft of an essay of letter.
53. I am good at thinking up puns.
54. I never use mental pictures or images when trying to solve problems.
55. While I have often seen pictures of him, I cannot remember exactly what President Reagan looks like.
56. I often remember work I have studied by imagining the page on which it is written.
57. Studying the use and meaning of words has become a habit with me.
58. I speak or write what comes into my head without worrying greatly about my choice of words.
59. Not enough people pay attention to the manner in which they express themselves.
60. I enjoy solving crossword puzzles and other games.
61. I find it difficult to form a mental picture of anything.
62. Memorizing by verbal repetition is time consuming and inefficient.
63. My dreams are extremely vivid.
64. I have better than average fluency in using words.
65. I read a great deal
66. I am continually aware of sentence structure.
67. My thinking often consists of mental pictures or images.
68. I do not form a mental picture of people or places when reading of them.
69. I often have difficulty in explaining things to others.
70. My daydreams are rather indistinct and hazy.
71. I find it easier to learn from a demonstration than from written instructions.
72. I often enjoy the use of mental pictures to reminisce.
73. I often use mental images or pictures to help me remember things.
74. When remembering a scene, I use verbal descriptions rather than mental pictures
75. I take great pains to express myself with precision and accuracy in both verbal speech and written work.
76. I have never done well in learning languages.
77. The proper use of words is secondary to the ideas and content of speech or writing.
78. I have a better memory for things I have read, rather than things I have experienced.
79. I am disturbed about people who quibble about word usage.
80. I have difficulty producing associations for words.

81. I often have ideas that I have trouble expressing in words.
82. I think that puns are the lowest form of humor.
83. Just before falling asleep I often find myself picturing events that have happened.
84. I prefer to read instructions about how to do something rather that have someone show me.
85. I am a good story teller.
86. I spend very little time attempting to increase my vocabulary.

Appendix D
Kirton Adaption-Innovation Inventory

1. Has original ideas
2. Proliferates ideas
3. Is stimulating
4. Copes with several new ideas at the same time
5. Will always think of something when stuck
6. Would sooner create than improve
7. Has fresh perspectives on old problems
8. Often risks doing things differently
9. Likes to vary set routines at a moment's notice
10. Prefers to work on one problem at a time
11. Can stand out in disagreement against group
12. Needs the stimulation of frequent change
13. Prefers changes to occur gradually
14. Is thorough
15. Masters all details painstakingly
16. Is methodical and systematic
17. Enjoys detailed work
18. Is (not) a steady plodder
19. Is consistent
20. Imposes strict order on matters within own control
21. Fits readily into "the system"
22. Conforms
23. Readily agrees with the team at work
24. Never seeks to bend or break the rules
25. Never acts without proper authority
26. Is prudent when dealing with authority
27. Likes the protection of precise instructions
28. Is predictable
29. Prefers colleagues who never "rock the boat"
30. Likes bosses and work patterns which are consistent
31. Works without deviation in a prescribed way
32. Holds back ideas until obviously needed