SANDIA REPORT

SAND2012-8590
Unlimited Release
Printed October 2012

Visual Inspection: A Review of the Literature

Judi E. See

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online order:	http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2012-8590
Unlimited Release
Printed October 2012

Visual Inspection: A Review of the Literature

Judi E. See
Human Factors and Statistics
Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185-MS0830

Abstract

This report provides a review of 212 documents published in the visual inspection literature from the 1950s to the present. The inspection task is defined and characterized and trends in inspection research are reviewed. The prevailing two-component model of inspection and various techniques to measure inspection performance are discussed. The majority of the report reviews the parameters that may impact inspection performance and describes the relevant findings from the literature. The report concludes with a re-iteration of the value of inspection research and a summary of the major recommendations for improving inspection performance.

ACKNOWLEDGMENTS

This report was completed under a Nuclear Safety Research and Development project funded by the U.S. Department of Energy, National Nuclear Security Administration. The report was developed in partial fulfillment of the final independent project for the Weapon Intern Program Class of 2012. Dr. Caren Wenner of the Human Factors and Statistics Department at Sandia National Laboratories in Albuquerque, New Mexico, provided management oversight and advice throughout completion of the project.

CONTENTS

FIGURES

TABLES

NOMENCLATURE

ASSIST	Automated System for Self-Instruction for Specialized Training
CA	correct accept
dB	decibel
dBA	decibel, A scale
EFT	Embedded Figures Test
EPAULETS	Enhanced Perceptual Anti-terrorism Universal Luggage Examination Training System
FA	false alarm
GAITS	General Aviation Inspection Training System
H	hit
HIT	Harris Inspection Test
IBM	International Business Machines Corporation
M	miss
NDI	nondestructive inspection
NSE	Nuclear Security Enterprise
SPL	sound pressure level
SRK	skills-rules-knowledge
SUNY	State University of New York
TSD	theory of signal detection
WAIS	Wechsler Adult Intelligence Scale

1. INTRODUCTION

This report provides a review of the visual inspection literature from the 1950s to the present. Although various reviews have appeared in book chapters and journal articles over the years, the findings from the literature may not be reported such that Nuclear Security Enterprise (NSE) employees can immediately understand and apply the concepts to nuclear weapons inspection work. Further, some reviews have focused on only one or two inspection topics (e.g., individual differences in inspection performance) and do not provide comprehensive summaries of the literature as a whole. Drury (1992) published the last comprehensive review of the visual inspection literature in the second edition of the *Handbook of Industrial Engineering*. The corresponding chapter in the third edition of the book published in 2007 is limited to the topics of human factors and automation in test and inspection. The current paper aims to extend the review to the present and provide a document that is readily available and useable for NSE personnel.

1.1. Approach

Two hundred twelve inspection-related documents written between 1958 and 2012 were reviewed and summarized during the course of the literature search. Of these documents, 140 were available in existing archives in the Human Factors and Statistics Department at Sandia National Laboratories in New Mexico. These documents were published between 1958 and 2004. The remaining 72 documents published 2004 or later were located by conducting four different keyword searches via Google Scholar:

- *Visual inspection* and *human factors*
- *Visual inspection* and *defect detection*
- *Visual inspection* and *precision manufacturing*
- *Multiple inspections*

1.2. Literature Search Characterization

The majority of the documents included in the current review were derived from the 1990s and 2000s (Table 1). Nine researchers were responsible for 57% of the literature (Table 2).

Table 1. Inspection Document Counts by Decade

Decade	Count
1950s	1
1960s	13
1970s	29
1980s	24
1990s	64
2000s	69
2010s	12
Total	212

Table 2. Primary Inspection Researchers

Author	Location	Count
Colin Drury	SUNY at Buffalo	55
Anand Gramopadhye	Clemson University, SC	24
Brian Melloy	Clemson University, SC	7
Sajay Sadasivan	Clemson University, SC	7
Prasad Prabhu	SUNY at Buffalo	6
Mao-Jiun Wang	Taiwan	6
Shannon Bowling	Clemson University, SC	5
Mohammad Khasawneh	SUNY at Binghamton	5
Adrian Schwaninger	U. of Zurich, Switzerland	5

By far, the most prolific and well-known researcher in the field is Colin Drury, a distinguished professor emeritus of industrial and systems engineering at the State University of New York (SUNY) at Buffalo. The second most productive author, Anand Gramopadhye, was a student of Dr. Drury's who later established his own inspection research laboratory at Clemson University in South Carolina.

Seventy-seven percent of the literature surveyed was published in the United States or the United Kingdom (Figure 1). Japan published 4% of the literature, while Hong Kong, Taiwan, Switzerland, and Ireland each published 2%. Countries in the "other" category included Australia, Canada, Chile, Germany, Israel, Italy, Malaysia, the Netherlands, Puerto Rico, Saudi Arabia, South Africa, and Thailand.

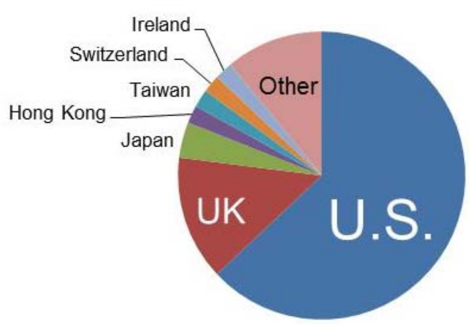


Figure 1. Categorization of Literature by Location.

Of the documents published in the United States, the majority originated from research entities in the northeast and southeast (Figure 2). The major contributors in these two regions were SUNY at Buffalo and Clemson University in South Carolina. Seventy nine percent of the documents published in the United States were accomplished through university research, most often conducted in an industrial engineering department. The remaining 21% were completed through corporations such as IBM or government organizations such as the Federal Aviation Administration. Some of the research in the published literature also originated from work conducted at Sandia National Laboratories, constituting a portion of the 9% of the literature from the southwest U.S. in Figure 2. For example, Alan Swain conducted research on human error rates and pioneered techniques in human reliability analysis in the 1960s and 1970s. Sandia National Laboratories also led an inspection research program in the 1980s and 1990s as part of the aging fleet evaluation program for the Federal Aviation Administration.

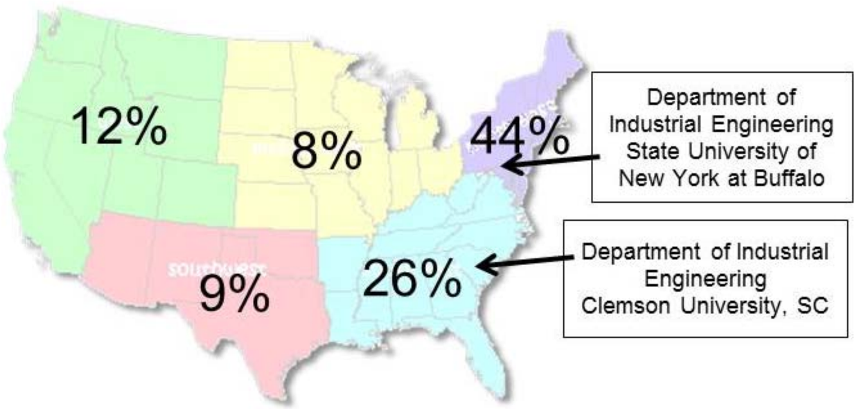


Figure 2. Percentage of Literature by U.S. Region.

1.3. Organization of this Report

The remainder of this report describes the technical results from the 212 documents that were reviewed. The review begins with a definition and characterization of the inspection task, a justification for researching the topic of visual inspection, and a review of trends in inspection research over the years. Drury's two-component model of inspection and various techniques to measure inspection performance are discussed. The bulk of the report reviews the parameters that may impact inspection performance and describes the relevant findings from the literature. The report concludes with a re-iteration of the value of inspection research and a summary of the major recommendations for improving inspection performance.

This Page Intentionally Left Blank

2. CHARACTERIZATION OF VISUAL INSPECTION

The dictionary definition of *inspection* is a *careful and critical examination, especially for flaws*. The term itself derives from the Latin *inspectionem*, meaning to *examine carefully* or *look closely into*. In fact, inspection is typically a deliberate, in-depth, exacting process that requires more than mere looking or scanning. As Drury and Prabhu (1992) point out; precision, depth, and validity are critical elements of the definition. Inspection processes require a large amount of mental processing, concentration, and information transmission, along with extensive use of both short-term and long-term memory (Gallwey, 1982). For example, short-term memory is required to remember which areas of an item have already been inspected and which have not. Long-term memory may be required to recall the standard dictating what a good item should like. In addition, inspection tasks are usually inherently stressful for inspectors. Inspection must typically be completed quickly, and critical defects that may require rework must be identified early. In addition, multiple defects at various severity levels and locations may be present, adding to the complexity of the task.

2.1. Inspection in Industry and Beyond

Inspection is a major quality control component for many industrial tasks. The intent of conducting inspection is to verify that a product is free of defects before being distributed for use. Inspection plays a role in many different fields, most of which are characterized as high-consequence:

- Food industry
- Aircraft maintenance
- Printed circuit assemblies
- Airport baggage screening
- Medicine (radiology, pharmaceuticals, histology)
- Nuclear weapons

For most of these fields, the inspection process is exacerbated by the fact that defects tend to be rare occurrences, while the costs of missing them is high. As one example, if tainted meat is missed during final inspection, people who consume the meat may fall seriously ill or die. Further, defects may be somewhat nebulous and difficult to define precisely. Airport baggage screening provides a good example of the challenges involved in defining “defects.” The catalog of potential threats changes constantly for airport baggage screeners. While the general nature of the threat is known (e.g., guns, knives, incendiaries); the actual size, shape, and characteristics change over time. Threats may also be deliberately concealed. The field of radiology provides another example of the difficulties inherent in inspection. For example, the incidence of breast cancer is very low, rare signs of potential cancer must be detected at their earliest presentation (which means detecting extremely small image changes), and cancer in the breast image may be masked by image background density (which varies with a woman’s age and physiology).

2.2. Importance of Inspection

The criticality of inspection in manufacturing and production becomes evident when the potential consequences of missed defects are examined. In some cases, missed flaws can have serious consequences ranging from injury to fatality. The field of aviation maintenance and inspection provides three examples:

- In 1988, an Aloha Airlines Boeing 737 lost part of its upper fuselage after cracks that were missed during inspection failed while the plane was in the air—one fatality and 65 injuries occurred.
- In 1989, a crash landing in Sioux City, Iowa, took place after a cracked engine fan disc that went undetected during inspection failed completely—111 fatalities occurred.
- In 1996, an engine failed during takeoff at the Pensacola, Florida, airport after a crack in the engine's front compressor fan hub was not detected during inspection. There were two fatalities and one serious injury.

In other instances, inspection errors may not cause injuries or fatalities, but they can translate directly into costs for the company. On the one hand, a defective product may be shipped, which may negatively impact customer satisfaction and reduce the likelihood of repeat business. On the other hand, a good item may be classified as defective and have to be reworked or scrapped, resulting in unnecessary expenses for materials and labor. As just one example, a Malaysian factory that produced printed circuit assemblies launched an investigation in 2004 to resolve inspection errors because too many defective products were being distributed and subsequently returned. The distribution of defective products degraded customer satisfaction and led to an annual loss of nearly $300k.

Finally, as in the case of nuclear weapons work, an inspection error could conceivably lead to both types of consequences. A flaw in the system that goes undetected might lead to a situation in which the weapon system detonates unintentionally, causing serious injuries or fatalities. In some instances, if radioactive material is released, the injuries or fatalities may not be immediate. Such an incident would also incur steep costs for the Department of Energy in terms of lost materials, medical expenses, investigations, and recovery.

2.3. Trends in Visual Inspection Research

Major trends in inspection research since the 1950s are identified by decade in Table 3. Such trends have paralleled the changes in manufacturing that have occurred since the inception of the industrial revolution. Early in the 20th century, industrial processes relied heavily on post-production inspection for quality control. A fully manufactured product was submitted for inspection and then either distributed to the customer base if it was judged to be acceptable or sent back to manufacturing for rework or scrap if it was judged to be defective. By contrast, modern production processes tend to de-emphasize or even criticize inspection as unproductive or counterproductive. Increasingly, the goal is zero-defect manufacturing, with an emphasis on in-process inspection and quality control at the source rather than the end of the process. In addition, the use of automation in both manufacturing and inspection has been on the rise in recent years.

The path that inspection research took began with the assumption in the early part of the 20th century that the outcomes of the inspection process were completely reliable. When it began to appear that this might not be the case, there was a concerted effort in the 1950s to study the inspection process in depth in order to characterize it. This trend continued into the 1960s with attempts to measure inspector performance accuracy and understand the inspector's thought processes during task completion by conducting detailed observations and interviews, an approach necessitated by the fact that inspection is a perceptual and cognitive process that cannot be observed externally. The theory of signal detection began to be applied to characterize inspection performance in the 1970s, and there were rigorous attempts to mathematically model the visual search phase of the inspection process. The 1980s saw a focus on understanding individual differences in inspection performance and attempting to devise techniques to select the best people for the job. The first study of automation in inspection also occurred in the 1980s (Drury & Sinclair, 1983). After the Aloha Airlines accident in 1988, there was an explosion of research in the 1990s to investigate the reliability of aircraft maintenance and inspection. As personal computers became more common in the workplace, the 1990s also saw numerous studies of the effectiveness of computer-aided instruction. The application of inspection lessons learned to airport baggage screening tasks began to occur in earnest after 9/11. The most recent inspection studies published 2010 or later tend to investigate the utility of virtual reality training techniques and automated tools for inspection.

Table 3. Trends in Visual Inspection Research by Decade

Decade	Research Focus
1950s	• Characterize inspection process • Dispel myths about inspection
1960s	• Conduct detailed observation and interviews • Measure inspector performance accuracy
1970s	• Apply signal detection theory to inspection performance • Mathematically model visual search phase of inspection
1980s	• Conduct initial explorations of automation • Study individual differences and develop selection techniques
1990s	• Investigate reliability in aircraft inspection • Study efficacy of computer-aided instruction
2000s	• Expand inspection research to other domains such as airport baggage screening
2010s	• Develop virtual reality techniques for training • Characterize utility of automated inspection tools

One observation has occurred repeatedly and consistently since the initial investigations of the inspection process—human inspectors are imperfect. As Drury (1992) stated, inspection error is a fact of life. It can be reduced with appropriate interventions, but it cannot be eliminated. After conducting a comprehensive review of 20 years of literature, Swain and Guttmann (1983) estimated that the minimum error rate appears to be on the order of 10^{-3}. This value applies primarily to relatively simple accept/reject inspection tasks. For more complex tasks, the error rate generally exceeds 1 in 1000 opportunities. In fact, as shown in Table 4, error rates of 20% to 30% are frequently quoted in the inspection literature across multiple types of inspection tasks (Drury & Fox, 1975). Further, 100% inspection does not alleviate the problem. Even under 100% inspection, not all of the defects will be detected (Drury, Karwan, & Vanderwarker, 1986).

Table 4. Observed Inspection Error Rates

Study	Task	Defect Detection
Lawshe and Tiffin (1945)	Measuring dimensions	9% - 64%
Hayes (1950)	Surface defects in piston rings	67%
Jacobson (1952)	Inspection of soldering defects	45% - 100%
Carter (1957)	Fabrication, coating, cutting, drilling, or beveling of acoustical tiles	76%
Heida (1989)	Magnetic particle inspection of aircraft main landing gear components	57% - 98%
Drury et al. (1997)	Aircraft visual inspection	68%
Leach and Morris (1998)	Close visual inspection of subsea structures and pipelines	53%
Graybeal et al. (2002)	Routine inspection of highway bridges	52%

Further examination of inspection errors reveals that they are more likely to be omissions (missing a defect) rather than commissive errors (false alarms—flagging a good item as defective) (Wiener, 1984). In addition, omissions are most likely to occur as a result of deficiencies when searching or scanning the item for defects. Omissions may occur for many reasons, but can be traced to task, environmental, individual, organizational, or social factors. Examples for each of these categories are provided below; however, it is important to note that omissions generally result from a combination of factors and not just one.

- Task: the task may be paced so quickly that the inspector does not have sufficient time to thoroughly inspect each item and therefore misses defects simply because they were never visually fixated in the first place
- Environmental: the amount of time spent continuously on inspection may induce vigilance effects that lead to lapses in attention and missed defects
- Individual: the inspector may lack a physical capability needed to perform the job (e.g., normal color vision) or have certain biases based on recent experience with defect types, rates, and locations
- Organization: training on the types of defects that might be encountered or the standards used for decision-making may be insufficient
- Social: implicit pressure from manufacturing to reduce the number of returned products may lead to “flinching”—accepting items that are borderline but should still be classified as defective products

2.4. Inspection Process

The process of visual inspection is essentially an ill-structured task with no simple step-by-step recipe for success (Drury, 1999). Typically, the knowledge base required of the inspector is large, while the indications of defects are visually difficult to discern (Drury & Lock, 1992). There are many nuances to performing an inspection task, including the fact that visual inspection is only partially defined by the visual sense. For example, the sense of touch may be used to augment vision when inspecting for surface defects such as roughness. The olfactory sense may be used to help detect fluid leaks or overheated control pivots during aircraft maintenance and inspection.

In general, there are five basic steps that must be completed during an inspection task:

1. Set up: obtain procedures or work instructions, items to be inspected, and equipment needed for inspection
2. Present: prepare item for inspection (e.g., by installing in holding fixture or microscope)
3. Search: examine item for possible defects
4. Decide: compare potential defect against standards to determine if it exceed the standards
5. Respond: accept or reject the item based on the decision made in Step 4, mark the item as needed, and complete required paperwork

The most widely recognized model of inspection is Drury's two-component model. The two-component model focuses on the *search* and *decide* components, which tend to be the most complex and error-prone steps in the process. These two components are independent processes that occur serially, with a finite time spent on the item to be inspected. In essence, the inspector must look for many things at once (*search*) and then accept or reject quickly (*decide*).

Spitz and Drury (1978) demonstrated the validity of the two-stage model's assumptions of independence and additivity in an experiment that involved a visual search task, a decision-making task, and an inspection task. The independence assumption implies that the variance in the inspection task times should equal the sum of the variances of the search and decision-making task times. Experimental results upheld this assumption. The additivity assumption implies that the decision-making task does not begin until the search phase has been completed. Experimental results indicated mean inspection times equaled the sum of search and decision-making times, providing support for the additivity assumption. Finally, the experiment demonstrated large correlations between predicted and actual inspection performance, as measured by the probability of a correct response as a function of response time (r = 0.88 or greater).

As further evidence of the independence of search and decision making, a study of inter-individual correlations in a variety of inspection tasks demonstrated that a decision-making factor was generalizable across different tasks, whereas the search factor did not correlate between tasks (Drury & Wang, 1986). This finding implies that if an inspector uses an optimal search strategy to inspect a circuit board, this strategy may not carry over to a different type of inspection task. That is, the skills and abilities important for search in one inspection task will not necessarily be valid in another inspection task. On the other hand, a poor decision maker tends to be a poor decision maker, regardless of the specific task involved (though proper training can help improve decision making in inspection). Thus, the decision-making component of inspection appears to reflect common underlying skills and abilities that apply across tasks.

2.4.1. Search Component

The search stage of inspection is generally bottom-up, rapid, and global. Bottom-up processing relies on basic visual features that tend to attract visual attention. These include color, size, orientation, edges, corners, and blinking or flashing lights. Thus, the search process consists first of a pre-attentive phase in which the eye is drawn to any very salient features in the field of view. Next, the eye fixates at each pre-determined location in an extended search. Nearly all of the information gathered during the search phase occurs during the fixations. Fixations average about 300 milliseconds in duration and account for more than 90% of the search time. During each fixation, the inspector can detect defects in an area of the item around the center of the fixation called the visual lobe. The visual lobe is the useful field of view that defines the limit of peripheral sensitivity for a given defect.

The search phase terminates when a defect is identified, the amount of time allotted for search has elapsed, or the inspector determines that enough time has been spent looking for defects. Numerous visual search studies have consistently demonstrated that search times are exponentially distributed (Drury, 1978). Optimal stopping policy in search can be determined based on the probability of a defect being present, the cost of an inspector's time, and the cost of missing a defect. In other words, search stops when the inspector perceives that the "cost" of further searching exceeds the "reward" for success. Indeed, Baveja, Drury, Karwan, and Malon (1996) showed that the stopping policy was close to that predicted by an optimization model based on costs, payoffs, and probabilities. Tsao and Wang (1984) further demonstrated that stopping policy during self-paced inspection of computer-drawn geometric patterns conformed to a combined reinforcement/expectancy theory. Specifically, reinforcement effects were stronger just after detection of defect, as evidenced by shorter stopping times, whereas expectancy effects became stronger later on, as evidenced by subsequent increases in stopping times.

Deficiencies in scanning technique may lead to poor defect detection during the search phase of inspection. For example, without training, people have a tendency to inspect the center of a display more closely than the edges. In general, search strategies during inspection can be random, systematic, or random with controlled overlap (Courtney & Guan, 1998).

- Random: fixations are chosen arbitrarily without regard to previous fixation locations; as a result, visual lobe overlap and repeated search of the same location may occur
- Systematic: the item is scanned methodically, as in reading, with little or no visual lobe overlap; systematic search implies perfect memory of previous locations fixated
- Random with controlled overlap: inspectors attempt to avoid refixating areas already searched, but some overlap occurs due to imperfections in memory

The type of search strategy adopted depends on the individual and task. The majority of subjects in Courtney and Guan's (1998) experiment used a random or random with controlled overlap search strategy. When inspecting color images of circuit boards, however, subjects adopted a systematic search strategy, starting at the top left and proceeding by "rows" to the bottom right of the board (Drury & Chi, 1995). A systematic search strategy such as this represents the most accurate and efficient strategy for most inspection tasks (Drury, 1990; Wang & Lin, 1993). In fact, training inspectors to use a systematic search strategy has been shown to significantly improve performance in terms of increased perceptual sensitivity for defects, fewer false alarms, and greater coverage of each item to be inspected (Watts, 2011).

2.4.2. Decision Component

The decision stage of inspection is more considered and involves both bottom-up (sensory) and top-down (cognitive) processes mediated by existing knowledge and expertise. The theory of signal detection (TSD) is most frequently used to model the decision process. TSD is a model of perceptual processing that provides a means to separate the effects of perceptual and nonperceptual factors in detection situations (Green & Swets, 1966). In short, TSD maintains that a defect may be missed because the inspector's perceptual ability to detect that defect was low or because the inspector adopted a conservative response bias (with an inclination to classify most items as good and only a few, with extremely large and salient flaws, as defective).

2.4.2.1. TSD Distribution of Sensory Effects

According to TSD, the signal or defect to be detected does not appear in isolation, but rather occurs against a background of noise that may stem from one or more sources. Part of the noise is inherent in the sensory process, emanating from the spontaneous firing of the nervous system, while additional noise may arise from changes in the environment or the equipment used to generate the stimuli (e.g., dust on the item to be inspected). Because this noise is always present in a detection situation, the observer's level of sensation will be greater than zero at any moment in time. Consequently, as a result of the presence of noise, the task is not simply to determine whether a signal is present or not, but instead to decide whether the magnitude of sensation for a given event is more likely to be the result of noise alone (no defect) or of signal-plus-noise (defect present).

Under the classic parametric model of TSD, the sensory effects produced by noise and signal-plus-noise are assumed to follow normal distributions with unit variance (Tanner & Swets, 1954). The mean level of excitation produced by the noise distribution depends upon the background intensity and will be greater than zero since noise is omnipresent. The addition of a signal to the background of noise shifts the level of excitation upward so that the mean of the signal-plus-noise distribution depends upon both background and signal intensity. Although the introduction of a signal increases the magnitude of sensation, it leaves the nature of the distribution of sensory effects unchanged.

2.4.2.2. Decision Rules and Response Outcomes

In deciding whether the magnitude of sensory stimulation is more representative of noise or signal-plus-noise, the observer is essentially faced with the task of testing a statistical hypothesis. The observer is assumed under TSD to estimate the probability that an observation arose from signal-plus-noise versus the probability that it arose from noise and to compute a "likelihood ratio" of the two probabilities. As the likelihood ratio increases in magnitude, the subjective odds that the stimulus came from the signal-plus-noise distribution become progressively greater. Furthermore, the observer is assumed to establish some critical value so that when the likelihood ratio exceeds this cutoff, the decision is that a signal has been presented. A likelihood ratio that falls below the critical level is assumed to be due to the presence of noise alone. As shown in Table 5, the decision for any given event can result in one of four outcomes. With respect to inspection, a hit constitutes correct detection of a defect, whereas a false alarm consists of incorrectly identifying a good item as defective.

Table 5. Four Possible Decision Outcomes in a Detection Situation

Response	**Actual Condition**	
	Defective (Signal-Plus-Noise)	**Good Product (Noise)**
Reject	Hit (H)	False Alarm (FA)
Accept	Miss (M)	Correct Accept (CA)

The critical value or cutoff point set by the observer represents the response criterion, and its location will affect the relative frequency of the four possible outcomes (Figure 3). The observer who establishes a high cutoff for the likelihood ratio is said to be cautious or conservative since the magnitude of sensation must be very high before the individual will decide that a signal has been presented against the background of noise. An inspector of this type is less inclined to decide "reject" and more likely to report "accept" on a given event. As a result, placement of the criterion upward along the sensory continuum toward the signal-plus-noise distribution results in a decrease in both hits and false alarms as well as an increase in both correct accepts and misses. When the criterion is located midway between the means of the two distributions at their intersection, the inspector exhibits no bias toward reporting either "reject" or "accept." For this neutral observer, the proportions of hits and correct accepts will be equal as will the proportions of false alarms and misses. On the other hand, when the inspector sets a low critical value for the likelihood ratio so that the criterion is shifted downward along the sensory continuum, the response criterion is said to be lenient. This individual requires only a minimal level of sensory excitation to decide that a stimulus came from the signal-plus-noise distribution. Because this observer is more likely to report "reject" than "accept" on any given trial, both hits and false alarms increase while both correct accepts and misses decrease.

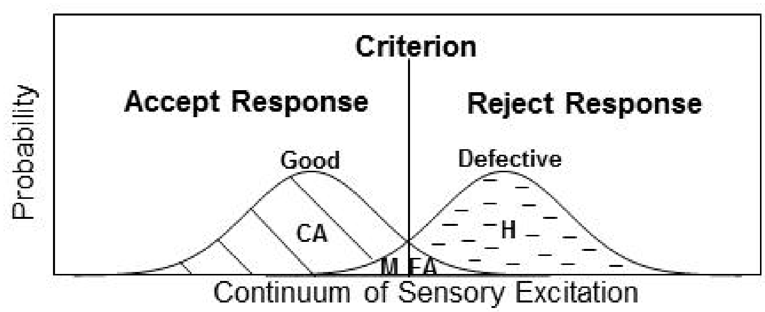


Figure 3. Response Criterion Impacts

2.4.2.3. TSD Performance Indices

The hit and false alarm responses made during a detection event are used to derive two independent performance measures: a measure of perceptual sensitivity and a measure of the response criterion. Both TSD indices can be calculated using only the proportions of hits and false alarms since the remaining two values are merely their complements (i.e., the proportion of misses is equal to 1 - H, and the proportion of correct accepts is equal to 1 – FA). The index of perceptual sensitivity is a perceptual measure that reflects the observer's ability to discriminate signals from noise, while the response criterion is a nonperceptual index that reflects bias in responding, or the observer's willingness to respond "signal."

The fact that TSD yields two independent measures of detection performance is perhaps its greatest advantage since it permits performance to be characterized in terms of both sensing abilities and decision-making processes. The two indices are assumed to measure different aspects of performance and to be controlled by different factors. The index of perceptual sensitivity is assumed to be affected only by the sensitivity of the perceptual system to the stimuli for detection, which in turn is affected by such perceptual factors as the salience of the signal to be detected. The response criterion, on the other hand, is affected by nonperceptual factors, including the observer's detection goals, expectations about the nature of the stimuli, the probability of signal occurrence, and the anticipated consequences of correct and incorrect responses.

Several alternative indices to measure sensitivity and bias are available. Parametric measures are used if the assumptions of normality and equal variance are met, whereas nonparametric indices are used if the assumptions are violated. Only the parametric measures are discussed here. The traditional parametric measure of perceptual sensitivity is the index d'. This index is essentially a standardized measure of the extent of the separation between the means of the noise and signal-plus-noise distributions. Graphically, the distance between the means of the two distributions grows larger as the observer's sensitivity increases, and the resulting value of d' will increase. General "rule-of-thumb" guidelines can be used to interpret the level of difficulty of a given detection task (Craig, 1984):

- Very difficult: $d' < 1.5$
- Moderately difficult: $d' = 1.5 - 2.5$
- Moderately easy: $d' = 2.5 - 3.5$
- Very easy: $d' > 3.5$

With respect to response bias, two parametric measures are commonly used. The traditional measure β represents the ratio of the ordinate of the signal-plus-noise distribution to the height of the noise distribution at the point in the sensory continuum where the observer's response criterion is located. β of 1.00 signifies a neutral or unbiased observer, β greater than 1.00 indicates a conservative criterion, and a value between 0.00 and 1.00 indicates a lenient criterion. The alternative parametric response bias measure, *c*, locates the criterion by its standardized distance from the intersection of the two distributions. The intersection defines the point where bias is neutral, and location of the criterion at that point yields a *c* value of 0. Conservative criteria yield positive *c* values, whereas liberal criteria produce negative *c* values. Multiple studies conducted in the areas of recognition memory and sustained attention have consistently demonstrated that the measure *c* is superior to β in its ability to reflect variations in response bias (Macmillan & Creelman, 1990; See, Warm, Dember, & Howe, 1997; Snodgrass & Corwin, 1988).

2.4.2.4. Application of TSD to Inspection

TSD provides a useful mechanism to explain many observations of inspection performance. For example, one of the primary findings in the literature is the imperfect performance associated with inspection. TSD by its very nature implies that 100% defect detection can rarely be expected and only at the expense of an increase in false alarms (unless the noise and signal-plus-noise distributions are very far apart). TSD also assumes that false alarms are an intrinsic part of any detection activity. Thus, in the case of inspection, it should be expected that some good products will be flagged as defective.

Characterizing inspection performance in terms of TSD can also provide useful information for management because it can help pinpoint the types of changes needed to improve performance. For example, if perceptual sensitivity is low, changes to enhance the salience of potential defects may be needed. Such changes might entail modified lighting, improved training, or the use of standards to promote comparative decision making. On the other hand, if response bias is too lenient or too conservative, simple changes in instructions may be effective. For instance, if too many defects are missed, inspectors could be briefed on the seriousness of letting defective products slip through the process (e.g., poor customer satisfaction, degraded product performance, and possible accident scenarios) and provided feedback on the current level of missed defects.

Drury and Addison (1973) demonstrated the value of TSD for inspection of glass items in an industrial study of the effects of feedback and defect density at Pilkington Brothers Ltd. Overall, the inspection group as a whole followed the theoretical predictions of the TSD model. For example, introducing more rapid performance feedback increased the effective detectability of defects significantly and reduced missed faults by half. This change was reflected in larger values of d', suggesting that inspectors' perceptual sensitivity was enhanced by providing rapid feedback on the accuracy of their inspection decisions. Average d' increased from 2.5 to 3.2 over the course of the 23-week study. According to Craig's rules-of-thumb, the initially moderately difficult task shifted to moderately easy after feedback was provided. Further, as implied by TSD, inspectors adjusted their response criteria to make fewer "reject" responses as the quality of the product improved (and thus presented a lower probability of defects). While inspectors adjusted their response criteria in the manner predicted by TSD, they also seemed to attempt to maintain the outgoing percentage of products judged defective at a constant level.

Drury (1974) further points out the utility of TSD for exploring the effects of the amount of time available to inspect each item. Namely, increasing the amount of time per item has a small but beneficial effect on perceptual sensitivity, but a drastic effect on the response criterion. As more time becomes available, the criterion is shifted lower and lower along the sensory continuum, resulting in more "rejects."

2.4.3. SRK Framework

The skills-rules-knowledge (SRK) framework developed by Rasmussen (1983) is useful for understanding behavior during both the search and decision-making components of inspection. This framework defines three types of behavior that describe how people extract, process, and understand information from the environment. The SRK framework represents a hierarchy of the level of processing needed to make sense of the environment.

- Skill-based level: information is perceived as continuous, time-space *signals* without any symbolic meaning. This level is the least effortful and essentially occurs unconsciously.
- Rule-based level: information is seen as *signs* that help the individual select or modify certain rules learned through past experience.
- Knowledge-based level: information is seen as *symbols* that can be used for reasoning and to generate new rules. This level is the most effortful.

The search component of inspection and the concomitant scanning and fixating behaviors are primarily skill-based. Relatively few knowledge-based behaviors are involved during search since there is less problem solving or active reasoning and more detection of defects. The decision-making component is dominated by rule-based and knowledge-based behaviors. Rule-based behavior is used during identification and classification of defects and application of rules or standards governing what to look for. Knowledge-based behaviors such as reasoning and deciding tend to be slow, error prone, and workload intensive. Once a certain level of expertise is gained, knowledge-based behavior is needed only when unfamiliar work situations are encountered.

Drury and Prabhu (1996) point out that the SRK framework is useful for understanding how behavior changes qualitatively as the inspector progresses from a novice to an expert. For example, because of reduced familiarity with defect detection, novices may identify some defects at the more effortful rule-based and knowledge-based levels, whereas experts can identify them at the less intensive skill-based level. The implication is that different types of information are needed to support novice and expert inspectors.

2.5. Measuring Inspection Performance

According to Johnson and Funke (1980), lack of information about inspector performance accuracy is one of the weakest links in establishing a quality control system. Actual levels of inspector performance are required before quantitative approaches to plan and control quality can be developed. For example, theoretically optimum inspection sampling procedures can be adopted, based on existing models, once inspector performance accuracy is known. Measures of inspection performance are also needed to effectively guide future performance and provide tailored training. In many cases, inspectors receive little or no feedback regarding the accuracy of their inspection decisions. If they do receive feedback, it is often far removed in time from the initial inspection. As a result, inspectors have no information upon which to base changes in their approach. That is, since they do not know if their decision is correct, they have no reason to change their approach. If inspectors are provided with data regarding their performance accuracy, however, they can use the information to understand and learn from their mistakes and modify future inspections. Finally, performance measures can identify deficiencies that may be resolved with tailored training. As one example, if an inspector consistently misses Defect X, training focused on helping the inspector recognize this defect and differentiate it from good products could be provided.

2.5.1. Alternative Measures of Inspection Performance

Numerous methods to measure inspector performance are available, and each has its advantages and disadvantages (Johnson & Funke, 1980; Sinclair, 1979). Sinclair (1979) categorizes the measures into four different groups:

- Empirical sensitivity measures: measures derived from quality control that have evolved to meet particular needs in industry
- Theory-based measures: measures that have arisen mainly as a result of TSD
- Speed measures: measures concerned with the relationship between the inspector's decisions and the time available to make them
- Measures of classification skill: measures concerned with the inspector's ability to classify inspected items correctly

The primary measures included in each group are described further in Appendix A. Advantages and disadvantages for each measure are also identified.

2.5.2. Criteria for Selecting Performance Measures

Performance measures should be selected first and foremost based on the organization's primary goals. For example, if the main objective is to complete inspection quickly so as not to delay subsequent production, one or more of the speed measures would be the preferred index of performance. Using a measure that matches organizational goals also facilitates development of appropriate training and job aids adapted specifically to meet those goals. In general, an index that is easy to understand is usually preferred over more complex measures. The measure should further be capable of differentiating among various levels of inspector performance. The selected measure should also be useful for identifying where the inspection process causes a drop in quality and no longer adds value.

Johnson and Funke (1980) concluded that the A_1 measure of proportion correct and the TSD measure of d' appear to be better measures of inspector detection efficiency than the alternatives in Tables A-1 through A-4 (Appendix A). One disadvantage of using proportion correct is the fact that it does not incorporate a corresponding measure of inspector response bias such as the TSD parameter β or c. Inspector response bias may not be an issue in cases where the relative costs of misses and false alarms are approximately equal, making the proportion correct an acceptable measure of inspector detection capability in these situations. Proportion correct is easier to compute than d' and does not require assumptions of normality and equal variance. In many inspection tasks, however, the cost of missing a defect is much greater than the cost of reworking a good product (false alarm). For these tasks, the TSD measures of d' and β or c would be the preferred measures. Finally, since neither A_1 nor d' incorporates measures of speed, it may be desirable to include one of the speed measures listed in Appendix A to supplement accuracy measurements.

2.5.3. Techniques to Measure Inspector Performance

Traditionally, there are three basic methods for measuring inspector performance:

- Conducting a separate test: a sample is developed that contains known good items and known defects; inspectors are given the sample as a test exercise and asked to find all the defects
- Labeling defects: known defects are marked inconspicuously, introduced into the process prior to inspection, and later retrieved from the *accept* and *reject* piles to measure performance
- Re-inspection: samples of items accepted and rejected by an inspector are re-inspected by another inspector to determine hits, misses, false alarms, and correct accepts

In recent years, computer-based techniques for simulating inspection have been implemented, primarily for aircraft inspection. Most of these techniques serve dual purposes to provide training and assess inspector performance. For example, Latorella et al. (1992) developed computer-simulated inspection tasks for nondestructive inspection (NDI) and visual inspection as a means to facilitate offline experimentation in aviation maintenance and inspection and to evaluate various models and interventions. In this case, the researchers did not set out to measure absolute values of defect detection probabilities or to train inspectors to accomplish the tasks involved. However, their tools can be used to provide indicators of inspector performance and deficiencies in the process. The Automated System for Self-Instruction for Specialized Training (ASSIST), on the other hand, was designed to provide a computer-based training tool for aircraft inspection (Gramopadhye, Melloy, Chen, & Bingham, 2000). The tool is intended to help improve the visual search and decision-making skills of aircraft inspectors by providing consistent training with feedback. However, like other computer-based tools, ASSIST can also be used to monitor and track inspector performance accuracy.

Regardless of which method is used to assess performance accuracy, the sample should include a representative range of defects, and there should be enough defects in each category to allow statistical analysis. The overall defect rate should also be representative of that typically found during inspection. A master list of defects against which to compare inspectors' responses must be developed and maintained. Finally, the sample should be altered periodically to prevent inspectors from recognizing defective items based on extraneous cues (e.g., scratches and markings).

In general, labeling defects provides the best indicator of everyday inspection performance since inspectors are not aware of being tested. Conducting a separate test (either with actual products or via computer-based techniques) provides an upper bound on performance since a test is usually not representative of normal working conditions. Finally, re-inspection provides a lower bound on inspection performance. The re-inspection process may be no better than the primary inspection, which can yield an inaccurate picture of detection performance. In fact, both inspectors may miss whole classes of defects.

This Page Intentionally Left Blank

3. PARAMETERS THAT IMPACT INSPECTION PERFORMANCE

The results of numerous inspection studies conducted over the past six decades have identified and clarified the factors that impact inspection performance. These factors fall into five different categories (Table 6).

Table 6. Factors that Impact Inspection Performance

Task	Individual
• Defect Rate • Defect Type • Defect Salience • Defect Location • Complexity • Standards • Pacing • Multiple Inspections • Overlays • Automation	• Gender • Age • Visual Acuity • Intelligence • Aptitude • Personality • Time in Job • Experience • Visual Lobe • Scanning Strategy • Biases
Environmental	**Organizational**
• Lighting • Noise • Temperature • Shift Duration • Time of Day • Vigilance • Workplace Design	• Management Support • Training • Retraining • Instructions • Feedforward Information • Feedback • Incentives • Job Rotation
Social	
• Pressure • Consultation	• Isolation • Communications

3.1. Task Factors

Task factors involved in inspection are associated with the physical nature of the inspection task. Task factors include features of the product to be inspected, the standards against which the product is compared, the manner in which the product is presented for inspection, and the availability of tools to assist in the inspection process.

3.1.1. Defect Rate

Defect rate refers to the probability of a defective product occurring in the batch of products to be inspected. In the majority of production situations, particularly in light of the modern emphasis on zero-defect manufacturing, this rate tends to be exceptionally low. Wiener (1984) indicated that defect rates of 1% to 10% are typical in most production tasks, referencing several studies of actual production that show defect rates ranging from 0.01% for atomic weapon components through 0.2% for coins and 4% for jam tarts. As a result, most inspectors may examine many hundreds or thousands of parts before they see a defect.

Multiple studies have demonstrated that inspection accuracy suffers as the defect rate decreases. That is, as the quality of the product improves, the likelihood of finding any defective products in the lot is reduced. In one of the earliest studies of defect rate, 80 inspectors completing a self-paced paper-and-pencil inspection task detected significantly fewer defects and committed significantly more false alarms as the defect rate decreased from 16% down to 4%, 1%, and .25% (Harris, 1968). A subsequent study involving detection of multiple defects in screws indicated that the defect rate may be confined to self-paced conditions only (Fox & Haslegrave, 1969). In that study, defect detection was poorer at a lower defect rate (.005 versus .05) only for inspectors in the self-paced condition who were able to manually select approximately 100 screws to inspect at a time. For those inspectors in a paced condition who had to inspect the screws as they passed by on a moving conveyor belt at 52 feet per minute, the defect rate effect was not observed.

Research by Brinkley (1994) further demonstrates that the defect rate effect may be most influential during the decision-making phase of the inspection process as opposed to the visual search phase. During no-decision visual inspection of critical soldering defects in printed circuit boards, performance was not significantly impacted by defect density level, which ranged from 5% to 15% defective boards.

Finally, the defect rate effect may disappear altogether if inspectors are allowed to correct their last decision. Fleck and Mitroff (2007) showed that detection was degraded only when observers in their study could not correct their last mistake. They maintain that rare “targets” are missed because observers adapt to the low probability by responding more quickly. In other words, misses are due to response execution errors (response bias in TSD terminology), not perceptual or identification errors (TSD perceptual sensitivity). When provided an opportunity to correct their last response, observers can catch their mistakes.

3.1.2. Defect Type

Defect type encompasses both the variety of potential defects in an inspection situation and the number of different types of defects that are possible. As described in Section 2.1, inspectors typically must search for many different defects at a time, some of which may be fairly ambiguous and difficult to define *a priori*. In general, inspection performance is degraded as the number and types of defects increases, primarily as a result of limitations of human memory (Dalton & Drury, 2004).

First, early research demonstrated that not all defects are created equal. In other words, accuracy in detecting one defect tends not to be related to accuracy in identifying another type of defect.

During inspection of tin plates for defects such as surface blemishes and non-uniformity of the coating, the average correlation among four different types of defects was only $r = 0.20$ (McCornack, 1961). Thus, a given inspector might be able to detect one type of defect with 100% accuracy, but perform very poorly against another type of defect. Similarly, Schoonard and Gould (1973) demonstrated that errors during inspection of integrated circuit chips tend to cluster on a few stimuli, rather than being uniformly distributed over all stimuli. Further, in another study of inspection of integrated circuit chips, it was observed that some defects were inspected just as accurately, regardless of the speed at which they were presented (Schoonard, Gould, & Miller, 1973). In this case, inspectors worked at their normal pace as well as 1.5, 2.0, and 3.0 times slower and twice as fast. Overall, doubling the normal rate of inspection increased misses from 23% to 30%, but accuracy was not degraded at higher speeds for some of the defects, primarily those defined as "easy." These latter two studies imply that a serial system in which each inspector looks for only a subset of defects might be beneficial for overall defect detection.

Second, inspector sensitivity to detect a particular defect is impacted by the number of other different defect types included in the inspection. In a simulated conveyor-paced task involving inspection of plastic blocks, discriminability significantly decreased as the number of defect types increased from one to two, three, four, or five (Ainsworth, 1982). The average d' score decreased from 2.8 with one defect to 1.6 with five defect types. In another study, when defects of different types were presented for inspection, the individually determined d' for each was higher than the corresponding d' obtained in the mixed presentation condition (Geyer, Patel, & Perry, 1979). In other words, if inspectors are tasked to look for only one defect at a time, their perceptual sensitivity for that defect is significantly higher than if they are tasked to look for that defect as well as others concurrently.

Megaw (1979b) observed that, regardless of the total number of potential defects, inspectors appear to search for a subset of about five. He points out that, without feedback, inspectors will not necessarily choose the most appropriate subset of defects. Along these lines, Rao et al. (2006) concluded that six defects may represent the practical maximum for an inspector. In accordance with research findings regarding defect type, Gallwey (1998) recommended searching for one type of defect everywhere on the product, then moving to the second defect type, rather than trying to search for all defect types concurrently in all areas of the product.

3.1.3. Defect Salience

Defect salience refers to the conspicuity of a defect, or its detectability against the background of noise in which it is embedded. Not surprisingly, inspection performance is typically higher for salient defects that are "easy" to detect. In one experimental investigation of defect salience, five subjects viewed slide transparencies that represented a good item, an easy defect, or a hard defect (Geyer, Patel, & Perry, 1979). A good item consisted of three circles equal in diameter. A defective item consisted of two standard circles and one deviant circle that was either smaller or larger in diameter than the standard. In the easy condition, this deviant circle was either 1.33 times larger or .83 times smaller than the standards circles. In the hard condition, the deviant circle was either 1.16 times larger or .90 times smaller than the standard circles. The difference in defect salience was reflected in inspector perceptual sensitivity scores. Average d' for the easy defect was 2.30, compared to 1.58 for the hard defect.

3.1.4. Defect Location

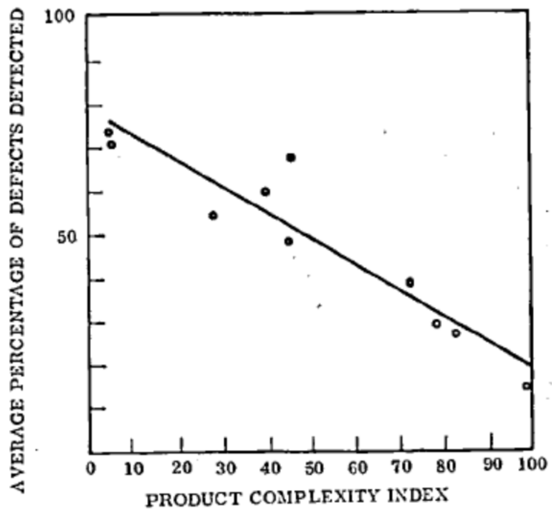
Investigations of defect location reveal that this factor may have little impact at a gross level, but more subtle impacts at a finer level of distinction. Namely, inspection performance for cases wherein defects are distributed across an entire item (e.g., aircraft fuselage) does not appear to differ from cases in which defects are confined to only part of an item (Gallwey & Drury, 1986). In that investigation, 66 subjects viewed transparencies consisting of computer-generated symbols to detect two, four, or six pre-defined defects. Distribution of defects varied such that all defect types were able to appear anywhere in the display or each defect type was confined to one side or the other of the item. Distribution of defects had no effect on performance.

Wilhelmsen, Ostrom, and Kanki (2002) demonstrated that the effects of defect location on inspection performance may be more subtle. Their study involved inspection of three different aircraft parts (panel, beam, and lug) for 16 crack lengths varying from 1/8” to 2”. Results indicated that cracks on parts without any reference points (e.g., the panel) appeared to influence inspector estimates of crack length more so than when a reference point was available (e.g., floor beam). For the floor beam, the presence of holes of standard size provided a reference point for inspectors. As a result, inspectors estimated crack length slightly more accurately for floor beams than for panels.

Finally, Kane, Moore, and Ghanbartehrani (2009) demonstrated that inspectors may develop a location bias, depending on prior experience. The task in their experiment was to inspect characters printed on a paper airplane for the defective letter *X*. For one set of paper airplanes, defects were distributed randomly over all parts of the plane. For a second set of paper airplanes, defects were randomly distributed on the wings only (location bias condition). After completing trials runs in either the unbiased or biased condition, subjects completed a similar inspection task in which defects were randomly distributed over the entire airplane. Subjects exposed to the location bias condition detected substantially fewer defects (36%) as compared to unbiased subjects (65%) outside the wings of the paper airplanes—the region where biased subjects had been exposed to low defect quantity in the trial runs. Thus, subjects developed a bias to search certain locations more thoroughly than others, based on their previous experience with the inspection task.

3.1.5. Product Complexity

Product complexity can be defined in different ways, depending upon the nature of the product to be inspected. In general, a complex item presents more parts or sub-components for inspection. As just a few examples, complexity for printed circuit boards has been defined in terms of the number of solder joints, whereas complexity for airport baggage screening has been defined in terms of the clutter within the luggage. The overall finding for complexity, across multiple fields, is that inspection performance declines as complexity increases. Harris (1969) reports a study in which equipment complexity was measured by counting the number of major parts comprising each item. The resulting complexity index ranged from 6 to 100. Results indicated that complexity has a significant negative effect on inspection performance (Figure 4). Ainsworth (1982) also showed that discriminability decreased significantly in a simulated conveyor-paced task as complexity increased from one to five sub-assemblies in a test item.



From Harris (1969)

Figure 4. Inspection Accuracy Declines as Complexity Increases.

Similar effects have been demonstrated when complexity is defined as background clutter. For example, Drury and Clement (1978) demonstrated that search time depends most heavily on the number of background characters in a display, as opposed to search field area or density of background characters. Number of background characters alone explained 86% of the variance in observed results.

3.1.6. Standards for Comparison

The manner in which defects are defined can impact inspection performance. Defect definition includes both the literal definition of the defect (e.g., a two-inch long scratch) and any standards that may be used to define it in relation to good products. First, with respect to literal defect definition, Jamieson (1966) indicated that the absence of a clear specification of what constitutes a defect contributes to poor inspection performance. Specifically, lack of clear definitions of defects prompted inspectors in the telecommunications industry to form personal criteria, which led to variability in work quality during inspection. Such subjective criteria appear to drift over time as well, such that products that would have been rejected at one time are accepted at another (Juran & Gryna, 1988). This phenomenon can be observed both between and within inspectors. Barring clear definitions, individual inspectors may actually revise their own work if, unknown to them, it is returned.

Second, the use of standards against which to compare a given item is generally beneficial. Using standards changes an absolute judgment task into a comparative judgment task, eliminating the need to rely on long-term memory of the standard. Therefore, inspection tasks that make use of standards for comparison tend to be associated with better defect detection (Jamieson, 1966). However, the standards themselves must be simple and easy to interpret. Gallwey and Drury (1986) showed that inspection performance was worse when different standards had to be used for different areas of the product as opposed to the condition in which the same standards for acceptance/rejection could be applied to any defect. In addition, a computer simulation task of a real-world contact lens inspection task revealed that search times, stopping times, mean inspection time, and decision-making performance per lens all worsened as defect standard complexity increased (Rao, Bowling, Khasawneh, Gramopadhye, & Melloy, 2006). For example, in that study, a *simple* standard indicated that light small or medium scratches on the lens were acceptable; large scratches were not acceptable. The comparable *complex* standard indicated that light small or medium scratches less than eight in number and up to a maximum of two in the optic zone of the lens were acceptable; heavy scratches or more than two light small or medium scratches in the optic zone were not acceptable. The authors concluded that inspection performance can be improved by using simpler defect inspection standards.

In general, standards should illustrate the differences between acceptable and unacceptable parts and should be easily accessible to the inspectors if they are to facilitate the task (McKenzie, 1958). However, such standards may not always be comprehensive or easy to obtain. For example, standard reference sets for visible particle contamination in pharmaceutical parenteral preparations are expensive to buy and do not represent all the different particles that could potentially contaminate pharmaceutical products (Sadeghipour, Bugmann, Herrera, & Bonnabry, 2007).

3.1.7. Pacing

Pacing refers to the amount of time available to complete the inspection task. In some cases, the rate of inspection is determined wholly by the system, as when conveyor belts are used to move products at a consistent rate and prevent bottlenecks. Conveyor-based systems are still widely used in the food industry. In other cases, the rate of inspection is determined by the inspector, within the limits set by the organization. During self-paced inspection, the inspector has the liberty to examine certain parts more thoroughly than others or to repeat an inspection if needed.

Overall, inspector-paced systems are typically more beneficial than externally-paced systems (Fox, 1973). The critical factor for externally-paced systems is whether the inspector's attention (i.e., decision to attend to the task or not) is synchronized with the appearance of a defect or even with the product itself. The optimal speed for conveyor-based systems depends on the product being inspected. For example, increasing the speed from 150 up to 200, 250, or 300 feet per minute during inspection of surface defects in sheet metal reduced inspection accuracy (Hayashi & Ogawara, 1977). At the higher speeds, only the more salient defects could be detected. Thus, in the case of sheet metal inspection, optimal conveyor speed appears to be approximately 150 feet per minute. For visual bottle inspection, on the other hand, the optimal speed is approximately 200 bottles per minute, which translates into a conveyor speed of 46 feet per minute (Saito & Tanaka, 1977). It should be noted that the speed imposed by externally-paced systems may impact more than inspection performance *per se*. For instance, at one factory implementing paced visual inspection of bottles, the absence rate increased as the work speed increased (Saito & Tanaka, 1977).

Generally, defect detection improves as the amount of time available increases, but at a decreasing rate and only up to a certain point (Drury & Watson, 2002; Megaw, 1979b). The net effect of slowing the rate of inspection is a 25% increase in the quality of inspection, but accuracy does not improve in proportion to the decrease in inspection rate (e.g., when the normal rate of inspection was doubled, misses increased from 23% to 30%) (Schoonard, Gould, & Miller, 1973). Further, previous research on visual inspection of integrated circuit chips has shown that performance eventually reaches an asymptote at about 120 seconds (Schoonard & Gould, 1973). In fact, in many instances, defects are usually detected quickly or not at all (Schoonard & Gould, 1973). One concern with increasing the amount of time for inspection is the false alarm rate. False alarms typically increase as the amount of time available for inspection increases (Fox, 1973). Drury (1974) pointed out that increasing the amount of time allowed per item actually has only a small albeit beneficial effect on detectability, whereas the effect on inspector response bias can be quite substantial. Inspectors tend to adopt a more lenient response criterion as the available time increases, such that they detect more defects but also commit more false alarms.

3.1.8. Multiple Inspections

Multiple inspections may occur within or between inspectors and can be beneficial if implemented correctly. A study conducted at the Autonetics Division of the North American Rockwell Corporation (now Rockwell International Corporation) revealed that repeated inspections significantly increased inspection accuracy for critical defects (at a slightly decreasing rate) up to a point (six independent inspections) (Harris, 1969). Little increase in inspection accuracy was noted when more than six independent inspections were used. In multiple inspector situations, the highest d' occurs when two inspectors inspect every item, and both must reject a product for it to be classified as defective (Drury, Karwan, & Vanderwarker, 1986). In that study, the researchers examined five possible methods to incorporate re-inspection with two inspectors:

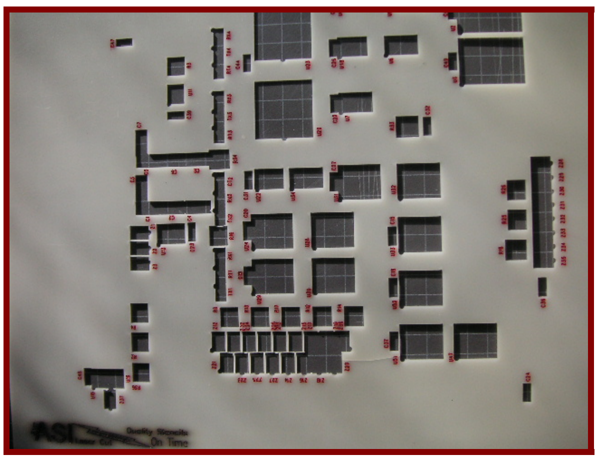
- Each inspector inspects only half the batch in parallel
- Two accepts: both inspectors inspect every item and both must accept an item for the system to accept it
- Two rejects: both inspectors inspect every item and both must reject an item for the system to reject it
- Re-inspect accepts: Inspector 2 inspects only those items accepted by Inspector 1
- Re-inspect rejects: Inspector 2 inspects only those items rejected by Inspector 1

The authors concluded that two inspectors are better than one in all cases except the parallel approach. Optimal detectability of defects occurs when the “two rejects” tactic is used.

3.1.9. Inspection Aids

Inspection aids can include magnification, directional lighting, image enhancement techniques, detailed requirements, computer storage of trend information and frequencies, and overlays (Gallwey, 1998). By definition, aids are designed to assist the inspector and facilitate the task, but there are limits to their effectiveness. Informing inspectors that a defect is present, for instance, does not actually aid performance since inspection accuracy does not increase (Schoonard & Gould, 1973). On the other hand, though, some very simple aids can have significant benefits. For example, inspection performance improved dramatically for 27 experienced machined-parts inspectors when visual aids were implemented (Chaney & Teel, 1967). The visual aids consisted of simple drawings of the parts to be inspected, with dimensions and tolerances for each feature. Use of visual aids alone resulted in a 42% increase in detection of objective defects (e.g., mislocated holes, threaded holes, and out-of-tolerance dimensions), without significant increase in inspection time or cost.

Overlays are materials that can be used to structure the product to be inspected. Figure 5 illustrates an overlay for printed circuit boards, complex products composed of a multitude of tiny components. The overlay has openings that reveal the parts of the circuit board that must be inspected. The rest of the overlay conceals the parts of the circuit board that do not require inspection. Research has shown that overlays can help structure the inspection process and improve defect detection by focusing the inspector's attention on the appropriate parts of the product (Fox, 1973). However, even a slight modification of the overlay approach may not benefit performance. In a study involving inspection of complex integrated circuit chips, for example, restricting the field of view to a series of small areas of the stimulus did not enhance defect detection (Schoonard & Gould, 1973). In this case, subjects serially viewed portions (one-fourth or one-sixteenth) of the product for a pre-determined amount of time, which is somewhat different from placing an overlay on the product, with all required inspection points displayed simultaneously.



From http://www.alt-solutions-inc.com/

Figure 5. Overlay for Printed Circuit Boards.

3.1.10. Automation and Inspection

Inspection has seen an increase in the use of automation since the 1990s. Before that time, the state of automation technology was not sophisticated enough to benefit the process. One of the earliest studies of automation for inspection occurred in 1983. Drury and Sinclair (1983) directly compared human and machine performance in a task involving inspection of 162 small steel cylinders. Four different types of defects were possible:

- Nicks and dents: elongated depressions in the metal
- Tool marks: gouges or cuts caused by one of the processing tools cutting too deeply into the metal
- Scratches: long, narrow damage marks
- Pits: small, circular depressions in the metal

Twelve human inspectors with at least two years of experience performed self-paced inspection of the cylinders. They worked at a slower pace than customary, inspecting the batch in four to six hours as compared to their normal pace of one to two hours. The prototype inspection device mechanically rotated the cylinder, performed optical scanning with a photo-diode array, and used a microprocessor-based detection circuit for defect detection. Results indicated humans performed significantly better than the machine, largely because of more sophisticated decision-making capabilities. The automated inspection device could locate most defects, but was unable to classify them as acceptable or not with the same consistency as the human inspectors. The machine had to operate at a very high threshold to approach human performance, which led to a high false alarm rate. The researchers concluded that more elaborate pattern recognition algorithms were required for machine performance to equal human performance.

Only one year after Drury and Sinclair's study, Wiener (1984) concluded that most defects are so ill-defined, subtle, and complex that no manufactured device could have the perceptual ability to recognize the defects and judge their acceptability. The picture had not improved much by the early 1990s. Automated inspection systems began finding increasing use in industry, but still had numerous limitations and deficiencies; they were also expensive, difficult to program, and extremely sensitive to normal variations in manufacturing (Drury & Prabhu, 1992).

A shift began to occur in the mid-1990s, however, as improvements in image representation and computational algorithms were realized. Automated devices have now become much more of a practical reality in many fields. For example, in a comparison of instrumented measurements of the flesh color and firmness of clingstone peaches with current subjective inspection methods, designed to sort the fruit into immature/mature and soft/firm categories, there was 83% agreement between the machine and human results (Slaughter, Crisosto, Hasey, & Thompson, 2006). The authors concluded that objective instrumental inspection methods hold promise as a replacement for current subjective methods. Similarly, an automated system to inspect automotive headlamp reflectors for defects that cause diffusion, distortion, or absorption of light showed promise during its development (Kuhlenkotter & Sdahl, 2007). Carrasco, Pizarro, and Mery (2010) describe an automated inspection prototype for image sequence acquisition and inspection of glass bottlenecks for defects. The prototype achieved 98% hits with only 2.4% false alarms during an experimental evaluation with 120 uncalibrated color images containing several hundred genuine defects.

Nevertheless, even with such successes, the use of automation in inspection must be approached with caution. For instance, issues of trust can be a critical factor in the successful implementation of automation. Operator trust in an automated device is based primarily on the perception of its competence (Kommidi, Dharwada, Gramopadhye, Cho, & Grimes, 2005). If the inspector does not trust the machine results, then any beneficial effects from automation will not be realized. In the Kommidi et al. (2005) study, a human inspector was required to supervise two computers and intervene as needed to override the automated decisions during simulated visual inspection of printed circuit boards. Three levels of automated competence were manipulated: both systems worked perfectly, both systems worked imperfectly, or one system worked perfectly and the other imperfectly. Results indicated that the competence of one automated system influenced human intervention behavior on the rest of the systems, leading to poorer performance. Another investigation of inspector trust in automation revealed that a machine aid with low misses and high false alarms yielded the best system performance and the best trust scores (Jiang, Gramopadhye, Melloy, & Grimes, 2007).

There is a fine line between an acceptable level of false alarms and too many, however. Too many false and nuisance alarms may erode the credibility of the alarms for the operator, who may then find creative ways to disable the alarms (Kraemer, Carayon, & Sanquist, 2009). For example, radiation portal monitors for cargo screening produce alarms from naturally occurring radioactive material in items such as ceramics, abrasives, and fertilizers (about 2% of screened containers). One explosive detection system for passenger air travel routinely operates at a 28% false alarm rate. Nuisance alarms may represent a small or large portion of all alarms that occur, depending on the frequency of occurrence of threats in each arena.

Finally, Drury and Watson (2001) caution that automation can yield reliable inspection only if the functions are allocated appropriately. Specifically, the parts of the job given to machines and to people must be appropriate to their different capabilities and needs. For example, in terms of the two primary phases of inspection, humans tend to perform better during decision-making as compared to search. Therefore, decision-making functions should be allocated to the human in the system, whereas search functions would be best allocated to the automated device. Additionally, humans are better than machines when dealing with novel situations, perceiving patterns, and forming generalizations based on individual occurrences. Machines tend to be better at performing repetitive, routine tasks and handling many complex operations simultaneously. As just one example, Drury and Watson (2001) recommended that the computer be allocated for measurement and calibration calculations during aircraft borescope inspection. A suitable function for the human during borescope inspection would include deciding whether or not to remove an engine that could legally be flown another leg despite the occurrence of defects.

3.2. Environmental Factors

Environmental factors that impact inspection performance involve the surroundings in which the work occurs. Environmental factors include physical components such as lighting, noise, temperature, and the design of the workplace. Environmental factors also stem from the manner in which the work is structured in terms of time of day and vigilance effects. Many environmental factors have not been widely investigated specifically within the field of inspection; hence, findings from other fields of study must be applied.

3.2.1. Lighting

The primary design requirements for lighting sources are brightness, glare, color, and focus. One of the chief difficulties with lighting for inspection work is achieving an appropriate brightness level without causing glare. Levels of 500, 1000, and 2000 lux are generally recommended for ordinary, difficult, and highly difficult inspection work (Megaw, 1979b). Given the detailed nature of inspection work, these levels are higher than those typically found in a normal office environment (300 to 500 lux). As Drury and Watson (2002) point out, vision can be improved by increasing the lighting level, but only up to a point, as the law of diminishing returns operates. Increased illumination can also result in glare.

Besides the overall level of lighting, the quality of the lighting is also important. Quality includes the color and focus of the lighting. With respect to color, a study involving inspection of sheets of steel was prompted by the fact that the factory began using green fluorescent lights in place of the normal white "daylight" tubes in order to conform to customer requirements (Dalton & Drury, 2004). Potential impacts of the change had not been evaluated beforehand. An offline experiment wherein the lighting could be controlled was conducted after the change and revealed that green versus white lighting did not significantly impact search time, hits, or false alarms during inspection. Inspectors committed more false alarms under white lighting (34%) versus green lighting (21%), though the effect was not statistically significant (possibly because the sample size of inspectors was only nine). In addition, during in-depth interviews after the experiment to explore inspection at the factory more thoroughly, the researchers discovered that inspectors who scored in the bottom half of the group in terms of performance tended to prefer white lighting. A search of the literature did not identify any further studies of the impacts of lighting color.

In terms of the focus of the lighting, some defects are more visible with predominantly diffuse lighting, while others are more salient with directional lighting. For example, during aircraft inspection, area illumination is needed when inspecting broad areas such as the fuselage for defects, whereas focused illumination is needed to complete finer inspection of a potential defect. One effort that specifically examined lighting for aircraft inspection involved installing a light shaping diffuser on an ordinary flashlight as a replacement to the lens (Shagam, Lerner, & Shie, 1995). The light shaping diffuser produced a more uniform beam of light that virtually eliminated glare. Field trials at multiple aircraft maintenance facilities demonstrated general acceptance of the light shaping diffuser, but primarily for close inspection work involving distances of less than 20 inches. A number of inspectors preferred the flashlight without the light shaping diffuser when performing general exterior walk-around inspections at night to view surfaces at distances of 20 feet or more. NDI inspectors also preferred flashlights without the light shaping diffuser since they use flashlights chiefly for area illumination while performing eddy current or ultrasonic inspections.

Achieving optimal illumination is not a trivial matter. Megaw (1979a) reported that a survey of quality control managers and chief inspectors of manufacturing industries in the United Kingdom indicated that nearly all companies had difficulties with lighting. As recently as the 1990s, Thackray (1992) discovered that illumination levels across 19 aircraft inspection and maintenance facilities were considerably below the recommended levels, partly because the lights were often placed too far away from the work being performed or because too few lights were used.

3.2.2. Noise

Findings regarding the effects of noise on inspection are derived primarily from the general human factors literature or from the field of vigilance. Overall, the general human factors literature indicates that, except for short-term memory tasks, the level of noise required to impact performance consistently is very high (95 dBA or more). Simple, routine tasks may exhibit no effect or even an improvement with noise. Detrimental effects of noise are usually associated with tasks performed continuously without rest pauses between responses and difficult tasks such as inspection that impose high perceptual or information processing capacity demands.

Conclusions derived from the field of vigilance may generalize well to inspection tasks because vigilance bears many similarities to inspection (vigilance is discussed in Section 3.2.5). Research has shown that vigilance efficiency may improve, decline, or remain stable in the presence of continuous noise (Hancock, 1984). The impact is dependent on the noise quality (white or varied), noise level (above or below 90 dB SPL), and degree of processing demand in the task (low or high). As just one example, low-intensity varied noise can facilitate vigilance performance for simple, low-demand monitoring tasks.

In the case of intermittent noise, performance improves when the presence or absence of noise is contingent upon the response. In one study, observers in a three-hour vigilance task received a broadcast from a local radio station based on their detection performance or heard the radio continuously, regardless of their performance. In the radio contingent condition, the radio was turned off after a missed signal and turned on when a signal was detected. Performance was significantly better in both radio conditions as compared to a control condition with no radio. However, observers in the radio contingent condition did not exhibit a performance decrement over time. As will be discussed in Section 3.4.4, this finding may have more to do with the knowledge of results provided in the radio contingent condition than with the presence of the background noise per se. Interestingly, Fox (1973) reports studies in which the presence of background music alone may improve defect detection by about 10% if used sparingly for short periods at prescribed intervals—in this case, the background music does not provide any feedback regarding response accuracy for observers.

3.2.3. Temperature

As with noise, a search of the inspection literature did not uncover any research that specifically examined the effects of temperature on inspection performance. Murgatroyd, Worrall, and Waites (1994) attempted to examine the impact of ambient temperature during inspection tasks in the aviation industry, but were unable to draw any conclusions from the limited data collected. Consequently, conclusions must again be drawn from other fields of study.

As in the case of noise, it appears that ambient temperature has the greatest impact on complex tasks as opposed to tasks involving simple mental activity or reaction time. For example, temperatures above 90°F tend to impair vigilance performance (Hancock, 1984). While cold temperatures have not been investigated as extensively, available evidence indicates that very cold conditions can also impair vigilance performance. Hancock (1984), for example, reports an experiment in which subjects exposed to cold exhibited longer response times as the watch period progressed.

Cold and hot temperature extremes are completely within the realm of possibility for both vigilance and inspection tasks. For example, lookout duties that require sustained attention can occur on ships operating in arctic conditions. Amusement ride inspectors may be tasked to inspect rides outdoors during the summer months, where temperatures may easily exceed 90°F (Woodcock, 2003). In many cases, such inspectors must wear long sleeves, long pants, and hard hats, which can add to the thermal load of the environment.

3.2.4. Time of Day

Many organizations involved in manufacturing and production incorporate multiple shifts that span 24 hours per day. Inspection is no exception. In fact, the majority of aircraft inspection is completed during the night shift between the last flight of the evening and the first flight of the next morning. The results of several studies suggest that this is the worst time of day for inspection:

- Drury (1974) reported that inspectors working night shifts are more likely to miss defects
- Wilhelmsen, Ostrom, and Kanki (2002) indicated that the swing shift appeared to be least accurate in estimating the length of cracks during aircraft inspections
- McCallum et al. (2005) demonstrated late night and early morning shifts for airport baggage screeners were associated with degraded attentional switching and poorer performance

3.2.5. Vigilance

Vigilance refers to the ability to sustain attention and remain focused on a task for prolonged periods of time in order to detect potential critical signals or targets. The systematic study of vigilance began during World War II when it was noted that airborne British radar observers began to miss the blips of light on their displays that indicated the presence of enemy submarines in the sea below after only about 30 minutes on watch. The Royal Air Force commissioned Norman H. Mackworth (1948) to study the problem in his laboratory. His investigations marked the beginning of over six decades of research of the *vigilance decrement*, or the decline in performance efficiency that occurs over time. The vigilance decrement has been replicated consistently since Mackworth's initial research, and it continues to be the most ubiquitous finding in vigilance experiments (Davies & Parasuraman, 1982; Warm, 1984). The decline in performance is typically complete 20 to 35 minutes into the session, and at least half of the final loss occurs during the initial l5 minutes of the watch (Teichner, 1974). Vigilance decrements tend to be worse for rare events, difficult detection tasks, situations in which no performance feedback is provided, highly repetitive tasks, and circumstances in which the observer is socially isolated during the task (Jerison, 1963; Nuechterlein, Parasuraman, and Jiang, 1983). As Dember and Warm (1979) pointed out, the most salient aspect of the vigilance decrement is that it seems to stem simply from the necessity of looking or listening for a relatively infrequent signal for a prolonged period of time.

Vigilance is frequently a component of the inspection process since inspectors must remain alert and attempt to detect potential defects for the duration of a shift. In fact, vigilance shares many characteristics of the typical inspection task—rare signals, prolonged time on task, high memory load, spatial and temporal uncertainty regarding signal occurrence, and the absence of performance feedback. Early studies of vigilance and inspection, however, concluded that vigilance and the vigilance decrement had little applicability for inspection tasks (Harris, 1969; Smith & Lucaccini, 1969). Such studies maintained that the vigilance decrement was an artifact of laboratory experimentation and the relatively simple tasks used for such studies, as compared to the complexities involved in most inspection tasks.

More recent studies, on the other hand, suggest that while defect salience and the complexity of decision making involved may be the primary determinants of inspection performance, vigilance can also play a role, depending on how the task is structured. In fact, the initial skepticism regarding the goodness-of-fit of vigilance for inspection appears to be largely unfounded because it was based on an early and outdated version of the vigilance model (Craig, 1983). Numerous real-world and laboratory studies of inspection have since demonstrated the impact of vigilance effects on inspection performance. Drury and Fox (1975), for example, noted that defect detection as a function of time could deteriorate up to 40% in 30 minutes, indicating that inspector performance cannot be completely described by decision and search models alone. Further, a study of inspection of rubber seals for automotive applications, which involved ten experienced inspectors working for 30-minute periods with a defect rate of about one per hundred seals, revealed a 27% decrement in hits from the first to the second 15-minute period (Fox, 1977). In addition, a study of professionally qualified aircraft inspectors working over a period of six days indicated that some cracks were missed due to inattention to the eddy current display (Murgatroyd, Worrall, & Waites, 1994). Drury and Watson (2002) also reported that several field studies of vigilance during non-aviation inspection tasks found some decrement in hits, ranging from 13% to 45%, associated with time on task. They judged that some vigilance decrement potentially exists for all inspection tasks. Finally, when 66 airline security screeners conducted an X-ray inspection task for four hours to detect eight threat types, a vigilance decrement occurred—detections and false alarms decreased from Hour 1 to Hour 4 (Ghylin, Drury, Batta, & Lin, 2007).

If vigilance is a factor in an inspection task, there are multiple recommendations to help prevent the occurrence of a vigilance decrement or reduce its magnitude. The principal recommendation is to alter the schedule so that inspectors spend no more than 30 minutes at a time inspecting parts (Drury & Watson, 2002; Fox, 1973; Gallwey, 1998; Purswell & Hoag, 1974; Saito & Tanaka, 1977; Swain, 1967). Purswell and Hoag (1974) advocated a five-minute rest break after 30 minutes of inspection work to relieve the perceptual and cognitive demands of inspection. They maintained that this work-rest schedule would be more beneficial than the traditional work-rest schedule of a 15-minute break during every four-hour period of work. Along these lines, Swain (1967) demonstrated that limiting the inspection period to 30 minutes reduced the number of missed defects during inspection of critical parts for military weapons, as compared to the current full-day schedule of inspection. The "rest" break does not have to be a non-working break, either. Simply alternating inspection with some other task approximately every 30 minutes is sufficient.

Other recommendations to combat the vigilance decrement during inspection include providing good feedback, social interaction, increasing the defect rate, and job rotation (Drury & Watson, 2002; Fox, 1973). Success has been achieved in some instances by inserting "false" defects to increase activity and defect detection. This approach has the effect of increasing the apparent defect rate and counteracting the negative effects of low signal probability. Providing opportunities for social interaction during inspection tasks can also help maintain performance over prolonged periods of time. For example, in a simulated paced inspection task, detection accuracy was greatest when inspectors worked in pairs, with each inspector selecting defective material from three lines of a six-line conveyor belt (Lion, Richardson, Weightman, & Browne, 1975). When inspectors worked alone at a single conveyor belt, they missed more defects and committed more false alarms. The authors concluded that *working in unison on a repetitive task stimulates performance.*

3.2.6. Workplace Design

A final environmental factor that can impact inspection is the design of the workplace in which inspection occurs. Namely, deficiencies in workplace design can contribute to poor inspection performance. As part of the aging fleet evaluation program conducted for the Federal Aviation Administration, for example, Thackray (1992) observed deficiencies in work support equipment across 19 aircraft inspection and maintenance facilities. Most facilities used a variety of fixed and moveable scaffolds and platforms, ladders, stools, and "cherry pickers." As the author pointed out, when work stands and platforms are not specifically configured for a particular aircraft, additional safety hazards may be incurred. Further, inspectors may experience increased fatigue if placement of the stands requires working in awkward positions. The inherent instability of cherry pickers, for instance, exacerbates difficulties associated with visual inspection and can be a source of distraction or concern for some inspectors, which can contribute to inefficient performance.

Yeow and Sen (2004) demonstrated that careful attention to proper workplace design can have significant payoffs in terms of inspection accuracy and customer satisfaction. In their study, ergonomics interventions were implemented to resolve three problems in a visual inspection process at a printed circuit assembly factory in Malaysia:

- Inspector eye problems (eye strain, headaches, watery eyes, and glare)
- Insufficient time for inspection (average of 7.5 components per second)
- Ineffective search strategy

The inspector problems manifested themselves in the large number of defective boards that were shipped to customer sites (approximately 2.7%). The defective boards were returned to the factory for repair and re-inspection, leading to an annual loss for the factory of approximately $300,000 and degrading customer satisfaction in the product. Ergonomics interventions reduced the percentage of defects at the customer site by 2.5% after just 12 weeks. One of the ergonomic issues rectified during the study involved glare from the templates used to inspect the circuit boards. Simply replacing the current kapton templates with less reflective white-colored paperboard templates and tilting the template to avoid specular reflection reduced glare by more than 35%.

3.3. Individual Factors

Individual factors refer to physical, mental, and personality characteristics of the inspector such as age, intelligence, and extraversion. Individual factors have been extensively investigated in an attempt to identify the traits of the "perfect" inspector and to develop personnel selection techniques for inspection. In fact, perhaps the most consistent finding in inspection is the existence of large differences both between and within inspectors, in the manner in which they perform the task and the overall level of accuracy they achieve (Drury & Wang, 1986). For example, in terms of differences between individuals, McCornack (1961) reported detection of solder defects for individual inspectors ranged from 43% to 100%, with no defect being detected by all inspectors. In terms of differences within individuals, McCornack (1961) also reported a study involving inspection of piston rings for surface defects in which 23% of the decisions were reversed when the rings were submitted twice for inspection.

During the first comprehensive review of research in individual and group differences in inspection, Wiener (1975) concluded that the general picture was *discouraging* and *inconclusive*. At that time, Wiener recommended that attempts to improve inspection and monitoring performance might best be served by looking elsewhere. The picture has improved since then, however, and the literature now provides evidence to support recommendations regarding the optimal tests to use for personnel selection. As Wang and Drury (1989) and Gallwey (1998) caution, though, no one test can be used for inspector selection for all possible jobs. As just one example, individual differences in inspection tend to be task-specific for the search component of the task, whereas decision tasks tend to show more generalizability (Drury & Wang, 1986). In the end, the very best predictor of inspection task performance is a simplified version of the task itself (Gallwey, 1982).

3.3.1. Gender

The traditional belief among manufacturing companies is that women are better suited to perform visual quality inspection than men (Heidl, Thumfart, Eitzinger, Lughofer, & Klement, 2010). This belief may arise from the stereotype that males should assemble items and females should inspect them or from the popular misconception that women are better at dull jobs (Wiener, 1975). Regardless of where it originated, this belief has not been supported by research. Wiener (1975), for example, reports 11 vigilance studies that examined gender. Six studies found no significant differences between the genders at all, two found that males were superior, and three reported non-significant main effects of gender but significant interactions with other variables. Similarly, Thackray (1993) reports that studies of individual differences in non-destructive inspection have not revealed any gender differences in task performance or in penchant for inspection. Heidl et al. (2010) did not find any overall differences between males and females who inspected 600 images of die-cast parts for defects such as scratches, cavities, and dirt.

3.3.2. Age

Studies of the effects of age on inspection performance have provided mixed results, though the findings are likely mediated by visual acuity and experience (Bloomfield, 1975; Fox, 1973). For example, some studies demonstrate that older inspectors have superior performance, but this effect may be due to their years of experience. Other studies demonstrate that younger inspectors have superior performance, but this effect may be due to their better visual acuity. Vigilance experiments have usually shown either no age effect or weak negative effects with aging. Czaja and Drury (1981) did find a significant effect of age in terms of longer search times for older inspectors, presumably due to a slowing in central perceptual processes that occurs with age, but no effect for errors. Wiener (1975) reported five studies with inconsistent results:

- Two studies reported a decline in detections with age and a significant negative correlation between d' and age, with d' declining about 0.2 units per 10 years of age
- One study showed the opposite correlation
- A fourth study found increasing accuracy up to age 34, then a decline to age 55
- A fifth study reported no age effect

If older inspectors do exhibit performance degradations, these may be overcome with routine training. Wales et al. (2009), for example, demonstrated that younger participants (average age of 21 years) achieved the highest hits and lowest false alarms in a simplified luggage inspection task. Overall detection ability as measured by the TSD index of perceptual sensitivity improved with training for both younger and older (average age of 60 years) participants; however, it took three training sessions for the older group to become as accurate and fast as the younger group was during its first session.

3.3.3. Visual Acuity

Visual acuity may be highly correlated with inspection accuracy or not correlated at all, depending on the sample involved. In a sample from the general population, for example, individuals with better visual acuity perform better on inspection tasks. In a sample from an existing inspector population, however, there tends to be no correlation between visual acuity and inspection performance, primarily because the individuals with poor eyesight have already been screened out. Thus, according to Drury and Watson (2002), inspector eyesight has never been found to be significantly related to performance for experienced inspectors. In fact, Wiener (1975) indicates that visual tests provide little guidance except to screen out personnel who lack the sensory capability to perform the task.

3.3.4. Intelligence

In his review of individual differences in inspection, Wiener (1975) concluded that intelligence did not provide a promising basis by which to select inspectors. Subsequent research, however, has shown that tests of intelligence may provide good predictors in some instances. In a simulated inspection task representing inspection of sheets of steel, Gallwey (1982) demonstrated that the attention-concentration subset of the Wechsler Adult Intelligence Scale (WAIS) can be a good predictor of inspector performance, *but only if* other better predictors such as the Harris Inspection Test (HIT) and the Embedded Figures Test (EFT) (covered in Section 3.3.5) are not used. The attention-concentration subset of the WAIS measures arithmetic, digit span, and digit symbol. A correlation between the WAIS attention-concentration subset and inspection performance supports the position that a large amount of mental processing is necessary for inspection work. Work by Wang and Drury (1989) further supports the notion that attention-concentration is a critical cognitive function for inspection work.

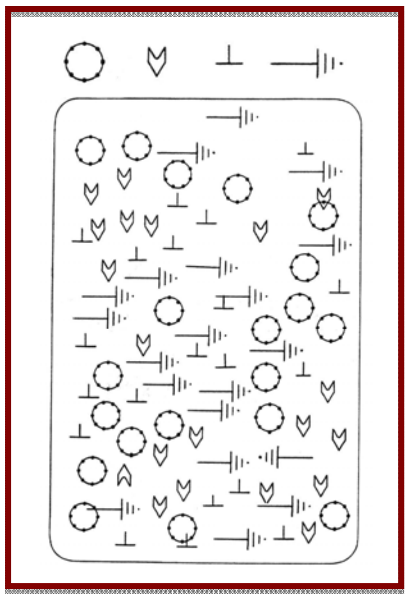
3.3.5. Aptitude

Aptitude tests measure various physical and mental competencies to perform certain types of work. Of the many aptitude tests that have been used to predict inspection performance, the two that have proven most beneficial are the HIT and the EFT (Drury & Watson, 2002).

3.3.5.1. Harris Inspection Test

The HIT is a paper-and-pencil test that can be administered in 10 to 20 minutes (Figure 6). Respondents are shown four different symbols whose configurations represent good products. They must then search a page containing approximately 75 symbols and designate any “defective” symbols that do not exactly match the standards provided. In essence, the HIT offers face validity as a measure of inspection performance since it is a mini inspection task.

The HIT has demonstrated significant correlations with inspection performance in many different inspection tasks. However, not all of the tasks that have been studied show significant correlations. Wiener (1975) reported significant correlations between the HIT and performance on three electronic inspection tasks. Correlations ranged from $r = 0.51$ to $r = 0.86$ in one study. Another study reported by Wiener (1975), on the other hand, showed negative results. In a study of search strategy training on a computer-simulated solder joint inspection task, the HIT predicted probability of detection (Wang, Lin, & Drury, 1997). Furthermore, following an experiment involving random search of 96 computer stimuli containing arrangements of simulated soldering points with one of three defect types, Wang and Lin (1993) recommended emphasizing the HIT as a predictor of error performance over the WAIS digit span test and the EFT.



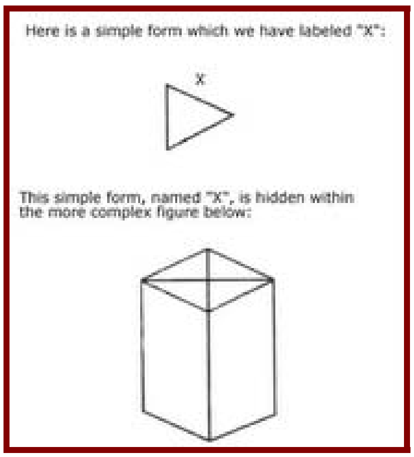
From Singh Negi (1981)

Figure 6. Harris Inspection Test.

3.3.5.2. Embedded Figures Test

The EFT is a standardized measure of cognitive style and analytical ability. The test requires finding simple forms that are embedded within larger figures (Figure 7). The EFT has 25 items that take approximately 20 minutes to complete. The score is the average time in seconds to detect the simple forms. Thus, higher scores indicate the individual has a tendency to perceive complete patterns rather than their separate components and therefore has greater difficulty analyzing a part separately from an overall pattern. Such individuals are classified as *field dependent*. Field dependent individuals have a more "gestalt" or holistic perceptual style and are more likely to attend to an item in its entirety rather than to its details. Personnel with lower scores are better able to separate components from the wider pattern and are classified as *field independent*. Field independent people tend to focus on details and to be more easily distracted by irrelevant cues.

Several studies in the literature have shown the predictive ability of the EFT for inspection accuracy. In Gallwey's (1982) study of individual differences in inspection using 10 different pre-tests, the EFT was a strong predictor of overall accuracy. The simulated inspection task, which was intended to represent inspection of sheets of steel, involved searching a 20 x 60 matrix of symbols containing up to six different types of imperfections. After finding a defect, subjects had to decide whether it was a second grade defect or a reject and then report verbally the judged size of the defect. In addition, of all the predictors examined in that study, the EFT was a predictor on the greatest number of measures of inspection performance—mean ending time, search errors, size errors, decision errors, and classification errors. The author does caution that the correlations obtained may have been inflated by the fact that the inspection task was geometrical in nature, and the EFT also uses geometrical patterns. Along these lines, in a similar study involving inspection of computer-drawn geometric patterns, field dependent subjects detected significantly more defects than field independent subjects (Tsao & Wang, 1984).



From http://aheadd.org/blog/wp-content/uploads/2009/06/embedded-figures-test1-272x300.gif

Figure 7. Embedded Figures Test.

The EFT has demonstrated applicability for non-geometric tasks as well as the ability to predict the search strategy inspectors will use. For example, the EFT predicted accuracy in detecting cracks on aircraft panels, although the correlations were not statistically significant (Drury, Spencer, & Schurman, 1997). Further, in a study of underwater close visual inspection and magnetic particle inspection of subsea structures, the correlation between the EFT and probability of detection was statistically significant (Leach & Morris, 1998). Finally, when 10 paid male students inspected color images of circuit boards for two defect types concurrently, subjects who searched for the easy defect first were all field dependent, whereas subjects who searched for both defects simultaneously were field independent (Drury & Chi, 1995).

3.3.6. Personality

When Wiener (1975) conducted his initial review of individual differences in inspection, personality tests had not been used for inspectors at that time. Based on the available evidence for vigilance or monitoring situations, he concluded that personality scales showed little promise for differentiating between good and poor monitors. In that review, Wiener (1975) reported an extensive study of multiple predictor variables for the performance of Navy personnel on visual and auditory vigilance tasks. Only 5 of 90 possible correlations with visual performance were statistically significant, and none of those predictors was a personality scale. Slightly more correlations with auditory vigilance performance were statistically significant (19 of 90), and three of those were personality variables. In general, introverts tend to be superior to extraverts during vigilance tasks. Although this finding is by no means clear-cut, no studies have shown the opposite. A search of the literature identified only one study since Wiener's review that specifically looked at personality scales and inspection. Thackray (1993) conducted an extensive study examining relationships among multiple aptitudes, traits, and performance on a simulated eddy current inspection task. Speed of inspection was positively related to measures of extroversion, impulsivity, and lack of meticulousness.

3.3.7. Experience and Time in Job

Common sense would suggest that inspectors with more experience and more time in the job would perform better than novices. Experienced inspectors have seen more defects and more varieties of defects than novices and have had more time to learn and remember the standards for comparison. The literature does not wholly support this supposition, however. McCornack (1961), for example reported that accuracy during tin plate inspection and during measurement of micrometer dimensions was not related to experience or length of time in the job. Correlations between accuracy and experience for tin plate inspection ranged from -0.07 to 0.06. In a more recent study of airport baggage screening, on the other hand, high screener performance was associated with more time on the job (McCallum, Bittner, Rubinstein, Brown, Richman, & Taylor, 2005). Unfortunately, the studies cited above did not report the ranges in experience that were examined.

A study of recognition of defects in chest X-rays did reveal some rather complex findings regarding previous experience and performance accuracy (Rebsamen, Boucheix, & Fayol, 2010). Namely, recognition rate for abnormal films increased in proportion to subject experience, but this tendency was reversed for normal films. This finding implies that experts may have lost their ability to recognize normal attributes. The same authors reached a slightly different conclusion in a study involving inspection of eyeglass lenses. Here they demonstrated that experts were able to detect the less extensive defects more quickly than novices, but there was no significant difference between the groups for very bad lenses. The authors concluded that inspectors gain the ability to recognize normal and abnormal configurations with experience, but only process the abnormal configurations. In this manner, the experts do not lose any time inspecting the normal lenses.

Finally, while experience may not be related to inspection accuracy per se, it may manifest itself in other ways. Megaw and Richardson (1979), for example, demonstrated that experienced or skilled inspectors searching tin cans for defects required fewer fixations, and their fixation times tended to be shorter. Rebsamen et al. (2010) also found that novices tended to more cautious (as measured by the TSD index of β) than experts during inspection of eyeglass lenses.

3.3.8. Visual Lobe

The visual lobe refers to the useful field of view, or the retinal field in which an object is capable of being noticed during a single eye pause. The visual lobe defines the limit of peripheral sensitivity for particular target and background characteristics. The size of the visual lobe is determined in part by the inspector's visual acuity and by the salience of the defect. In general, performance accuracy degrades and search time increases the further an item is located away from the fixation point and into the periphery (Drury & Clement, 1978). As a result, visual lobe size typically provides a good predictor of inspection performance (Gallwey, 1982). That is, inspectors with larger visual lobes tend to detect more defects or require less time for search.

Visual lobe size can be increased with training as one means to improve both the speed and accuracy of inspection. In one study, repeated practice in rapidly detecting a defect in peripheral vision in a computer-simulated visual inspection task increased visual lobe size, and this capability transferred to a more realistic visual search inspection task (Latorella, Gramopadhye, Prabhu, Smith, & Shanahan, 1992). Similarly, Gramopadhye and Madhani (2001) demonstrated that practice improves the visual lobe, which in turn improves visual search performance. They cautioned, however, that considerable practice is required to realize the full benefits of the training.

3.3.9. Scanning Strategy

As discussed in Section 2.4.1, scanning strategy is a critical factor in the search component of the inspection process. Research on inspector scanning strategy has revealed three major findings in terms of individual differences (Megaw & Richardson, 1979):

- Eye fixations tend to be shorter for easy tasks and for experienced or skilled searchers
- Experienced inspectors tend to require fewer fixations
- Accurate and efficient inspectors tend to adopt a systematic search strategy

Several studies have demonstrated the superiority of the systematic search strategy for inspection. For example, Wang and Lin (1993) showed that this strategy was the most accurate and efficient under various combinations of search field size and target type during inspection of soldering points of integrated circuit boards. Similarly, in a study by Drury (1990) involving stationary targets in a stationary field, systematic search always outperformed random search. Along these lines, Watts (2011) demonstrated that training inspectors to use a systematic search strategy significantly improved performance for 24 participants who inspected 90 parts from five different types of casting. Participants who underwent systematic search strategy exhibited increased perceptual sensitivity, fewer false alarms, and greater coverage of each casting.

3.3.10. Biases

Individuals may approach the inspection task with pre-conceived notions or biases that can impact performance. Experienced inspectors can build up expectancies regarding typical defects, locations of defects, and relative numbers of defective items. As a result, performance accuracy may improve with experience. However, experience may also work against the inspector. For example, if product quality changes gradually such that the defect rate subtly increases, the inspector may miss defects if he/she continues to operate under the assumption that the defect rate is still at the level previously experienced. The impact of such a bias was identified earlier in Section 3.1.4. Specifically, subjects exposed to a visual defect location bias during practice, wherein 90% of defects were distributed on the wings only of a paper airplane, detected far fewer defects outside the wings as compared to unbiased subjects during the actual task (Kane, Moore, & Ghanbartehrani, 2009).

Wiener (1984) identifies two inspector biases that are commonly observed:

- Censorship: excluding unacceptable findings
- Flinching: accepting borderline products that are only slightly defective, but still contain defects

Censorship and flinching can occur if an inspector receives pressure from manufacturing or management to accept as many products as possible, and some inspectors may be more susceptible than others. An example of censorship might be deliberately accepting defective products in order to remain within a tolerable reject rate, as implied by manufacturing or management. An example of flinching might be rounding off a measurement that lies between .758 and .759 to the higher value if that value brings the product within specifications. Here, the inspector gives the product the benefit of a doubtful reading.

3.4. Organizational Factors

Organizational factors refer to the larger structural, administrative, and political environment in which the inspection tasks occur. Such factors include management support, the amount and type of guidance provided for inspection, performance feedback, incentives for good performance, and job rotation.

3.4.1. Management Support

Management attitude toward inspection can impact performance effectiveness. If inspectors perceive that management does not value inspection, they may not be motivated to perform at their best. If inspection is viewed as a necessary evil, less skilled or less experienced employees may be assigned to the task.

A search of the literature identified only one study that directly addressed the effect of management attitude on detection performance. In that study, 112 Fort Knox armor trainees participated completed a three-hour vigilance task in order to detect random interruptions of a pilot lamp. The trainees were treated in either a democratic (expansive and permissive) or an autocratic (brief and brusque) manner by the experimenter. Under the democratic treatment, the experimenter explained the task requirements, demonstrated the task, provided the rationale for the research, and encouraged the trainees to ask questions. The experimenter also explained that different environmental conditions would be presented during the experiment and allowed the trainees to select the order in which they would encounter them. Under the autocratic treatment, on the other hand, the trainees were informed that various environmental conditions would be presented in a pre-determined order during the experiment, and any trainee questions were addressed in a curt manner. Results indicated that the trainees treated in a democratic manner exhibited significantly higher levels of detection performance than trainees treated in an autocratic manner.

Another set of studies did not directly address the impact of management attitude, but did look at a variation of this concept in the context of vigilance. Wiener (1975) reported three different studies in which instructions were varied to enhance or diminish the importance of the inspection task being performed. In one study, this difference involved changing only one word in the instructions—*challenging* versus *monotonous*. Subjects who received more positively toned instructions detected more signals.

3.4.2. Training and Retraining

Following a task analysis of inspection activities at representative general aviation facilities, Gramopadhye, Desai, Bowling, and Khasawneh (2003) identified potential interventions to minimize inspection errors. Training for inspection showed up most frequently as the intervention strategy of choice. Numerous studies in the literature have demonstrated the value of consistent and comprehensive training (though it is by no means a panacea for all ills). Training innovations do not require extensive changes to system hardware and are generally easy to implement. In addition, the magnitude of performance improvements that can be achieved with training is often greater than can be accomplished by other ergonomics interventions. As Gallwey (1998) pointed out, the benefits of a good training program can be very significant over the long term due to savings in unnecessary rework and scrap. Further, training can benefit both novice and experienced inspectors. For example, although ultrasonic inspectors working at the Ignalina Nuclear Power Plant in Lithuania initially performed well, performance was significantly improved with appropriate training that concentrated on their weak areas (Birchall, Worrall, Murgatroyd, & Saburov, 1994). In another study, a 32% increase in detection of objective defects (mislocated holes, threaded holes, and out-of-tolerance part dimensions) was achieved for 27 mature, experienced machined-parts inspectors following implementation of a four-hour training program (Chaney and Teel, 1967). Although the benefits of training have been well documented, it is a managerial decision to determine the type and amount of training to provide. For that reason, training is included as an organizational factor that can impact inspection.

3.4.2.1. Training Approaches

Proven techniques for implementing training have been identified on the basis of a vast body of literature on training in general as well as inspection-specific studies of training methods. Proven training techniques for inspection incorporate eight different features:

- **Use progressive-part training and a modular approach:** training parts of the task to criterion and then teaching successively larger sequences of the task produced large improvements in all aspects of performance during visual and tactile inspection of metal cylinders; repair and scrap rate at the manufacturing facility fell from 50% to 26% in the third month after training (Kleiner & Drury, 1993). A modular approach enables inspectors to acquire a range of knowledge and skills in an orderly manner and also facilitates the use of focused refresher training.

- **Provide rapid and reliable feedback during training:** feedback is an essential requirement for almost all types of instruction that promotes learning via its motivational effect, reinforcement of responses, and the informational content contained therein (Embrey, 1979). Feedback should be rapid and reliable during training to support development of a mental template of the characteristics that define defective and good products.

- **Use cueing to provide feedforward information:** cueing involves the presentation of a series of examples of defective and good products, each of which is classified as such before presentation; cueing tends to be equivalent to or superior to feedback during training (Embrey, 1979).

- **Employ active methods that require frequent trainee responses:** active training results in significantly better performance (fewer classification, search, and decision errors) as compared to training in which participants are passive observers (Czaja & Drury, 1981).

- **Incorporate self-paced discovery methods:** training is most beneficial if trainees are able to work at their own pace and have some responsibility for the direction the training takes. One approach involves using task cards, each of which specifies a block of behavior to be learned (Kleiner & Drury, 1993). Trainees read the information on the card aloud and then perform any required actions themselves, with guidance, cueing, and feedback from the instructor. Trainees decide when they are ready to proceed to the next card.

- **Clearly define acceptability standards and their application:** if the standards for accepting good products and rejecting defective products are poorly defined, inspectors form their own personal criteria, which leads to variability in inspector performance and product quality (Jamieson, 1966). During a study of quality control in the optical industry, a six-month training program was implemented to provide precise representations of categories of defects, the features of a normal eyeglass lens, and specific knowledge about acceptability standards and intensities of defects (Rebsamen, Boucheix, & Fayol, 2010). At the end of six months, novices and experts performed equally well when detecting good lenses. Although the experts maintained a slight advantage in overall errors (1.69% for experts versus 2.19% for novices), novice performance was vastly improved after the training.

- **Provide an error key of the most common errors:** a study that involved training image interpreters to identify targets in aerial photographs indicated that false alarms were reduced if the error key analyzed the most common errors from previous interpretations (as compared to a key that delineated characteristic features of the various targets) (Embrey, 1979).

- **Include a wide variety of defects:** inspectors are best prepared to detect novel targets when trained with diverse categories, as demonstrated in a simulated luggage screening task that manipulated diversity during training (Gonzalez & Madhavan, 2011). Subjects who saw targets from five of five possible categories during training had significantly more hits, fewer false alarms, and faster detection times during transfer than subjects who saw targets from only one of the five categories during training. Presenting a wide variety of defects permits inspectors to construct a mental model of the characteristics of good and defective products via experience with authentic examples. The defect rate presented during training may also need to be higher than the actual product defect rate. In an airline luggage screening task, participants trained at higher base rates of signal probability obtained higher hit rates at transfer, although they also had higher false alarms (Madhavan, Gonzalez, & Lacson, 2007). A moderate base rate of 50% during training had the most beneficial effects on detection sensitivity during transfer and a more optimal balance between hits and false alarms.

3.4.2.2. Range of Training

Both the procedural and the cognitive aspects of the inspection task can be trained using the approaches identified in Section 3.4.2.1. Procedural aspects of the inspection task include differentiating multiple types of defects from one another and from good products, using consistent names for defects, and applying standards for acceptance/rejection appropriately. Cognitive aspects of the task include applying a search strategy during inspection and deciding whether a product is good or defective.

As described in Section 3.3.9, a systematic search strategy is generally superior to a random search strategy. Research has demonstrated that training inspectors how to apply a systematic search strategy improves inspection performance (Wang, Lin, & Drury, 1997). In that study, participants who were trained in a systematic search strategy on a computer-simulated solder joint inspection task produced significantly higher performance in terms of both detection accuracy and search time than participants who used a natural search strategy. Random search training actually degraded performance. Another study demonstrated that the search performance of novice inspectors could be improved by providing cognitive feedforward training about the search strategy adopted by an expert inspector in a collaborative virtual environment (Mehta, Sadasivan, Greenstein, Gramopadhye, & Duchowski, 2005). During the training, novices were able to observe the expert's scanning pattern in real time via one of three different display techniques. The decaying trace display, which provides a brief positional history of the expert's sequence of eye movements, proved to be most effective, resulting in an increase in the number of defects detected. In this type of display, yellowish green dots faded away into the environment at an interval of 200 msec. The trail of dots representing the expert's point of regard while performing inspection was seen as a transparent line moving continuously, similar to a trace.

3.4.2.3. Training Tools

Tools used to facilitate training range from low to high fidelity, and all varieties can be beneficial if they incorporate the guidance recommended in Section 3.4.2.1. Traditionally, training has involved classroom presentation using paper-and-pencil methods, view graphs, and PowerPoint slides. More sophisticated techniques have emerged in recent years as the technology has advanced, leading to a proliferation of computer-aided training and virtual reality simulators for inspection.

Basic knowledge of training principles can be combined with advanced technology to design *computer-aided training* programs for inspection. Training principles provide the basic foundation for the training approach, whereas the advances in technology provide new methods for delivering the training. Studies in the 1990s, when computers began to be more common in the workplace, revealed that performance could be improved using computer-aided training. For example, studies in the Department of Defense revealed a 34% increase in outcome performance measures and 55% reduction in learning time with computer-aided instruction (Gramopadhye, Bhagwat, Kimbler, & Greenstein, 1998). One reason for this improvement was the fact that computer-aided training could be implemented more consistently than the traditional training in which inspectors were trained by different instructors, who each had their own individual approach. In a case study of contact lens inspection training conducted in the 1990s, 24 naïve subjects with minimal inspection experience completed either the existing training program or a computer-based progressive-part training program (Gramopadhye, Bhagwat, Kimbler, & Greenstein, 1998). Results indicated the computer group had significantly lower mean inspection times, search times, stopping times, and search errors per lens, with fewer defect classification errors and lens classification errors than the control group trained under the traditional program.

Since the 1990s, computer-aided training programs for inspection have grown ever more sophisticated. Such systems have become common in large industries characterized by very high throughputs, such as aircraft inspection and airline luggage screening training. Four such systems are described below.

- **GAITS:** the General Aviation Inspection Training System (GAITS) is a computer-based training system developed for aircraft inspection to standardize and systematize the inspection process in general aviation (Jacob, Raina, Regunath, Subramanian, & Gramopadhye, 2004). GAITS contains four main modules that provide information about various types of aviation inspection, specific training with quizzes on six different aspects of the inspection process, and a simulator that provides hands-on experience inspecting an aircraft part. GAITS training can be customized for individual training needs. At this time, the impact of GAITS training on subsequent inspection performance has not yet been documented in the literature.
- **ASSIST:** ASSIST is a computer-based training tool for aircraft inspection that is designed to help improve the visual search and decision-making skills of aircraft inspectors (Gramopadhye, Melloy, & Nickles, 2000). ASSIST was intended to systematize the inspection training process to overcome the drawbacks of existing on-the-job training, which is often characterized by infrequent or delayed feedback. ASSIST incorporates many of the recommended training principles, including pre-training, feedback, active training, progressive-part training, schema training, and feedforward training. The value of ASSIST was demonstrated in a study of 18 aircraft maintenance inspectors, half of whom received ASSIST training before completing an inspection task. The inspection task involved searching for seven computer-simulated airframe structural defects such as cracks and corrosion. The inspectors who received ASSIST training detected more defects and performed better under both unpaced and paced conditions.
- **EPAULETS:** the Enhanced Perceptual Anti-terrorism Universal Luggage Examination Training System (EPAULETS) is under development to improve performance in baggage screening, where errors tend to be due to faulty detection or interpretation rather than search per se (Gale, Purdy, & Wooding, 2005). EPAULETS uses eye movement recordings to quantitatively assess error performance. The impact of EPAULETS training on subsequent inspection performance has not yet been documented in the literature.
- **X-Ray Tutor:** the X-Ray Tutor is an adaptive computer-based training program designed around the factors that influence object recognition—viewpoint in which an object is depicted, superposition by other objects in the bag, and the number and type of other objects in the bag. X-Ray Tutor training begins with threat items presented in easy views and progresses in difficulty to more difficult views, more complex bags, and greater superposition. Two studies have demonstrated an increase in perceptual sensitivity and a reduction in screening time required following X-Ray Tutor training (Ghylin, Drury, & Schwaninger, 2006; Koller, Drury, & Schwaninger, 2009).

Virtual reality simulators provide an even greater level of sophistication than computer-aided training methods. As with the computer-based training techniques, they tend to be more common for aircraft inspection training than for other types of inspection. Virtual reality simulators provide immersive, interactive, three-dimensional computer-generated environments that seek to mimic the real world as closely as possible. In a direct comparison of a virtual reality simulator for aircraft inspection training and the ASSIST trainer, trainees preferred the virtual reality simulator (Vora, Nair, Gramopadhye, Duchowski, Melloy, & Kanki, 2002). Subjects felt that experiences in the simulator were as natural as real-world ones, and the virtual reality simulator scored higher than ASSIST in terms of its ability to depict defects. The virtual reality simulator was also associated with better inspection performance—simulator trainees detected significantly more defects with less visual search time. While the virtual reality simulator was viewed more positively overall, it was not considered to be as responsive to trainee actions as ASSIST. Further, a virtual reality system is more expensive and difficult to develop.

Two other studies led by Sajay Sadasivan have further demonstrated the value of virtual reality simulators for aircraft inspection training. The first study showed that a collaborative virtual reality training environment provided exposure to a wide variety of defects and produced a greater improvement in accuracy as compared to a control group (Sadasivan, Rele, Greenstein, Gramopadhye, Masters, & Duchowski, 2005). The second study demonstrated that virtual reality simulators do not have to have extremely high fidelity to be effective (Sadasivan, Vembar, Washburn, & Gramopadhye, 2007). This study revealed no significant differences in hits among three different interface conditions representing varying levels of fidelity. Further, subject perception of presence in the environment was equivalent across the three conditions. The results of this study suggest that lower-cost projector-based virtual reality simulators may provide a viable cost-effective alternative to the more expensive fully immersive virtual reality simulators.

3.4.3. Instructions and Feedforward Information

The amount and type of information provided during inspection is one of the most critical components in the process. The human inspector acts as an information processor during inspection, sensing information in the environment, comparing it to information stored in long-term memory, and making decisions on that basis (Drury, 1989). According to Thomas and Seaborne (1961), the inspector's task is almost entirely a function of the information received—instructions, messages, and consequences of previous performance. In fact, Drury (1989) maintains that information quantity and quality have the largest potential impact for both the search and decision-making components of inspection. Thus, it is no surprise that the design of information flow to and from the inspector impacts performance.

Information in inspection consists of one of three types:

- Directive: presentation of information in a form suitable for the human inspector (e.g., drawings optimized for inspection versus manufacturing)
- Feedforward: informs the inspector of the defect types and rates to expect
- Feedback: provides knowledge of results to indicate whether the correct acceptance/rejection decision was reached or the optimal search strategy used

A key contributor to inspector reliability is supplying the inspector with the appropriate information; i.e., deciding what information to present, when to present it, and how to present it. Performance can be affected by incomplete or insufficient information as well as too much information, exhibiting an inverted U function. Performance can also be affected by the quality of the information presented. For example, simple human errors can stem from inadequately defined inspection procedures. When the performance of professionally qualified inspectors working over a period of six days was assessed, considerable variability in calibrating ultrasonic inspection equipment was observed for most inspectors, even though the standard procedures were used (Murgatroyd, Worrall, & Waites, 1994). In addition, because the procedures did not specify the method of fixing a straight edge to a row of aircraft fasteners, one inspector used a handheld method that proved difficult at knee level. In fact, Drury and Watson (2002) reported a study indicating that 46% of all errors in maintenance and inspection had *documentation* as one contributing factor. In a case study of one inspection document, the error rate for instructions violating existing research guidelines for instructions was 1.4%. The corresponding error rate for instructions that complied with existing guidelines was 0%.

3.4.4. Feedback

As discussed in Section 3.4.2.1 on training, feedback is an essential requirement for nearly all types of learning because of its motivational and informational properties. Multiple studies in inspection and vigilance have demonstrated the value of feedback for performance effectiveness. In a study of inspection performance at the Pilkington Brothers Ltd. glass manufacturing facility, the introduction of more rapid feedback produced a large and consistent increase in d', reducing missed defects by half (Drury & Addison, 1973). In this case, 100% of the products were inspected immediately and classified as either good or defective. The 100% inspection was followed by a sample inspection by special examiners. To provide more rapid feedback to the 100% inspectors, the special examiners were moved to a point closely following the 100% inspection, and their results were made known to the 100% inspectors more quickly and directly than before. This change led to the observed increase in inspection performance during 100% inspection. Inspector d' scores continued to remain at this elevated level four months after the study was completed.

The efficacy of performance feedback and cognitive feedback was compared in a study using a realistic simulation of an aircraft structural inspection task (Gramopadhye, Drury, & Sharit, 1997).

- Performance feedback consists of feedback on performance metrics such as search time, search errors, and decision errors
- Cognitive feedback provides information on the process or strategy by which the inspector achieves certain performance results; cognitive feedback may be statistical (e.g., percentage of search area covered, number of fixations, and mean inter-fixation distance) or graphical (e.g., a visual representation of the scan pattern used)

Results indicated that performance feedback produced the greatest improvements in performance measures, doubling speed for almost no change in accuracy. Subjects who received performance feedback also adopted a more efficient search strategy, with fewer fixations and less overlap between fixations. Cognitive feedback enhanced the efficiency of the search strategy, with the graphical feedback producing the best combined response in performance and strategy. Subjects who received graphical feedback exhibited improved speed without affecting accuracy, and they had fewer fixations, shorter inter-fixation distances, and less overlap between fixations. Statistical feedback tended to increase speed, but at the expense of accuracy, making it less viable as an approach for providing feedback.

3.4.5. Incentives

Incentives for good inspection performance can take the form of financial rewards, verbal recognition, or offers of time off for high performance. Payoff matrices that reward correct responses or penalize errors are commonly used to study the impact of incentives on performance. The optimal response criterion for a given inspection situation can be calculated based on the probability of a defect and the rewards and penalties involved. Research has shown that people tend to modify their criterion in the direction of the optimum criterion based on the payoff matrix, but they are usually less extreme (Drury, 1978).

Results regarding the impact of payoff matrices on inspection performance have been mixed. Wiener (1975) reported that an offer of a *substantial financial reward* did not significantly influence inspector accuracy. He suggested that offering time off for high performance in place of bonus pay might be more motivating. In a more recent study of incentives for inspection performance, on the other hand, there were significant effects of rewards (Watanapa, Kaewkuekool, & Suksakulchai, 2012). The inspection task required subjects to detect errors in computer images containing Thai and Roman characters. Subjects either received no training, training with feedback, or training with feedback and monetary rewards for accurate performance. Results indicated that the group receiving the rewards missed the fewest number of defects. Thus, in this study, incentives positively impacted performance.

It should be noted that while payoff matrices are easy to implement in laboratory studies, they can be difficult to apply in actual industrial inspection, especially if they are based on performance accuracy. In real-world inspection situations, the accuracy of inspector decisions may not be evident for some time after the work has been completed. Further, payoffs can negatively influence inspector judgment in real-world tasks. For example, paying inspectors more for products that pass may lead to undesirable behaviors such as flinching or censorship.

3.4.6. Job Rotation

Section 3.2.5 discussed job rotation in terms of its ability to alleviate vigilance effects during inspection. It is included here since deciding whether to implement job rotation for inspectors is an organizational or managerial decision that can impact inspection performance. For example, as mentioned earlier, Swain (1967) showed that rotating inspectors to another task after 30 minutes of inspection reduced the number of missed defects during inspection of critical parts for military weapons, as compared to the current full-day schedule of inspection. In effect, job rotation serves to relieve the constant perceptual and cognitive demands of inspection.

3.5. Social Factors

Just as inspection occurs in the context of an organizational and political environment, it also involves an interpersonal and social milieu. Managers, manufacturers, and inspectors must work together to produce and distribute a completed product. Each of these sectors has slightly different goals, which may conflict at times and impact the inspection process.

3.5.1. Pressure

Inspectors may face pressure from multiple fronts. For example, they may be influenced by production personnel to accept their work because rejects can result in lost pay for manufacturing. Inspectors, particularly those who work in aircraft maintenance and inspection, are also generally pressed for a quick turnaround (Taylor, 1990). In the airline industry, inspection must frequently occur in between the last flight of the day and the first flight of the morning, pressuring inspectors to complete their tasks within that time frame or risk grounding a plane. Generally, aircraft inspector pressure is greatest at the beginning and end of this time period. When the aircraft arrives at the inspection maintenance site, the inspector is tasked with quickly identifying any critical defects that might require lengthy maintenance. At the end of this period, inspectors must quickly conduct "buy back" inspections to verify that the completed maintenance work was adequate.

In addition to these types of pressures, inspectors may receive pressure from management to minimize rejects since the consequences of shipping poor quality goods are far less immediate than the consequences of the rejection decision. Typically, when an item is rejected, additional paperwork must be completed and the item must be submitted for rework and re-inspection. To avoid that aggravation, inspectors may adopt a bias to accept most products. In fact, Wiener (1984) indicated that social and economic pressures push inspectors toward leniency (i.e., accepting borderline goods).

3.5.2. Isolation

Isolation refers to separation of inspection functions from manufacturing functions. McKenzie (1958) suggested that co-location of inspection and production may negatively impact inspection performance due to some of the issues described in the previous section. Namely, personal relationships between inspectors and manufacturers can subtly influence inspector judgments without their awareness. Along these lines, Jamieson (1966) indicated that efficiency in the visual inspection of telephone racks improved when inspection was isolated from production. Wiener (1975) also stated the quality of inspection may be higher when inspectors are spatially separated from production workers, but the same situation may not prevail for production workers—that is, production quality may increase when inspectors are nearby due to the immediacy of feedback on product quality for manufacturing. Wiener suggested that cross-training production employees and inspectors may be beneficial for the process as a whole.

An interesting study by Foot and Russon (1975) provides evidence, however, that such cross-training may introduce additional problems. In that experiment, members of 12 male subject pairings alternated in their roles as "inspector" and "operator" on a discrimination task. The operator performed a discrimination task involving size judgments of a series of colored comparison triangles or rectangles against a standard triangle or rectangle. The operator simply had to accept or reject each geometric shape on the basis of whether he judged it to be the same size as the standard or not. The inspector then had to evaluate the accuracy of the operator's responses, using either his own subjective judgment or by objectively matching each geometric shape against a template. In reality, the experimenter interceded so as to manipulate the nature of the feedback provided to operators, leading them to believe the inspectors had evaluated their performance either favorably or unfavorably on the basis of either subjective judgment or objective criteria. When their inspection discriminations were rejected by means of subjective rather than objective judgment, operators tended to retaliate. That is, the quality of feedback the subject gave when acting as an inspector reflected the quality of feedback he received when acting as the operator, but only if he thought this feedback represented personal opinions. If he received feedback that his decisions were wrong, he tended to reciprocate when serving as an inspector and flag the new operator's decisions as wrong. Thus, alternating inspector and manufacturing roles might set the stage for a tug of war between inspection and production.

3.5.3. Consultation and Communications

Consultation among inspectors relates to the degree of isolation they experience amongst themselves. In an early study, Thomas and Seaborne (1961) suggested that collaboration and communication over time may lead to increased consistency among inspectors. This supposition was confirmed in a study of consultation among radiologists regarding patient diagnosis based on X-ray film inspection (Hillman, Swensson, Hessel, Gerson, & Herman, 1976). Six groups of three radiologists each viewed abnormalities in 24 different X-rays and were asked to reach a diagnosis for each case. Three different conditions for diagnosis were involved:

- Without consultation
- With consultation leading to a diagnosis by group consensus
- With consultation followed by independent diagnoses

Consultation among radiologists led to overall performance improvements, regardless of whether consultation was followed by an individual or a group diagnosis. After consultation, 83% of participants improved their initial individual scores.

McCornack (1961) reported a study that demonstrated the powerful effect communications can have on inspector attitude and performance. In that study, inspectors were told that a set of rejected piston rings had been reworked prior to re-inspection. Operating under the belief the rings had been reworked, the inspectors accepted 67% of the previously rejected rings. When the inspectors were told the rings had not been reworked, on the other hand, 64% of the rings were again correctly classified as defective. One simple communication changed the nature of the task and had a drastic effect on performance.

This Page Intentionally Left Blank

4. SUMMARY AND CONCLUSIONS

Inspection is a major quality control component for many industrial tasks, and there is a vast body of research on visual inspection spanning the 1950s to the present. The impacts of numerous factors have been documented in the literature and summarized in this report. Inspection errors can have serious consequences, ranging from customer dissatisfaction and exorbitant costs for product rework to injury and death.

Perhaps the primary lessons learned from a review of the inspection literature include the following:

- Human inspectors are imperfect
- Large individual and group differences in performance exist
- Multiple differences in performing inspection tasks have been observed

4.1. Recommendations to Improve Inspection Performance

A multitude of solutions to improve inspection performance and overcome observed deficiencies have been identified and implemented in many real-world instances. In fact, Drury and Watson (2002) provide a very useful list of 58 good practices in visual inspection derived from industry sources and human factors analyses of visual inspection. This list not only identifies each good practice but also identifies the applicable phase in the inspection process (five steps described in Section 2.4) and provides a rationale explaining why the recommendation is considered a good practice. Potential solutions can be classified into three major categories.

- Training
- Inspection procedures
- Apparatus

4.1.1. Training

As discussed in this report, training provides a very effective means to improve inspection performance. Both the procedural and cognitive aspects of inspection are trainable, and substantial improvements in performance effectiveness can be achieved. Any training program should adhere to the proven principles summarized in this report. Periodic refresher training should also be implemented to improve the reliability of highly practiced inspection tasks. Refresher training can help combat problems associated with experienced inspectors who begin to perform the task from memory and may not adhere strictly to the official procedures.

4.1.2. Inspection Procedures

Revising the manner in which the inspection process occurs represents another avenue for improving inspection. The research reviewed in this report provides several suggestions:

- Inspect defect-by-defect rather than search for all defects simultaneously
- Train and use a systematic search strategy to maximize accuracy and minimize false alarms during search
- Incorporate up to about six independent inspections to increase accuracy
- Have two inspectors inspect every item and classify an item as defective only if both inspectors reject it

- Implement a serial approach in which each inspector looks for only a subset of defects (up to a maximum of six defects per inspector)
- Allow inspectors to correct their last decision to minimize the impact of response execution errors
- Reduce reliance on memory by using standards for comparison
- Minimize vigilance effects by incorporating recommendations identified in Section 3.2.5
- Consider using automation to help remove known defect-free items from the inspection system
- Provide feedback as often as practicably possible to provide knowledge of results regarding prior decisions

4.1.3. Apparatus

Apparatus includes the tools used directly to complete the inspection task (e.g., microscopes, magnifying glasses, and eddy current displays) and ancillary tools used to support or facilitate the task (e.g., fixed and portable lighting and equipment stands). Recommendations to improve inspection performance that involve apparatus include the following:

- Ensure equipment is properly calibrated before use as variability in calibration will translate into variability in inspection and product quality
- Provide lighting sufficient for the inspection task, as identified in Section 3.2.1
- Use overlays or templates to the extent possible to organize complex products and assist inspectors in conducting optimal search

4.2. Conclusion

Significant payoffs can be realized with relatively simple improvements in inspection procedures and apparatus. The study of visual inspection processes at the printed circuit assembly factory in Malaysia described in Section 3.2.6 provides a case in point (Yeow & Sen, 2004). In that study, ergonomic interventions reduced the percentage of defective boards arriving at customer rates from 2.7% to 0.2%, saved the factory over $250k per year, and improved customer satisfaction. These benefits were realized following several simple changes in procedures and apparatus:

- Reduced excessive use of a magnifying glass from 4.4 hours per day to 1.1 hours
- Replaced current kapton templates with less reflective templates and tilted the template to avoid specular reflection to reduce glare
- Limited visual inspection to electrically non-tested components only to reduce visual inspection from 7.5 components per second to 2.4 components per second
- Introduced a systematic visual inspection sequence via a revised circuit board template and instructions
- Advised leaders to fill in as needed and allow inspectors to take short breaks in order to relieve eye strain, headaches, and watery eyes

At the very least, simply attempting to define and obtain a test sample in order to assess inspector reliability can alleviate many inspection problems (Drury, 1992). When using this approach, Drury discovered that many problems arose because of inconsistencies in naming defects and interpreting standards. Further, defects seen by one inspector as occurring randomly were seen by others as occurring in runs. In essence, the simple act of asking the questions needed to collect a test sample solved numerous observed problems and improved the inspection process. Exhorting inspectors to try harder to be perfect is seldom the answer—it only increases the stress associated with the inherently stressful job of inspection (Drury, 1992).

This Page Intentionally Left Blank

5. REFERENCES

1. Ainsworth, L. (1982). An RSM investigation of defect rate and other variables which influence inspection. *Proceedings of the Human Factors Society Annual Meeting, 26,* 868-872.
2. Baveja, A., Drury, C.G., Karwan, M.H., & Malon, D.M. (1996). Derivation and test of an optimum overlapping-lobes model of visual search. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 26,* 161-168.
3. Birchall, P.D., Worrall, G.M., Murgatroyd, R.A., & Saburov, Y. (1994). *Evaluation of the initial effect of training on the capability of Ignalina inspectors*. Warrington, UK: AEA Technology.
4. Bloomfield, J.R. (1975). Studies on visual search. In C.G. Drury and J.G. Fox (Eds.), *Human Reliability in Quality Control* (pp. 31-43). London: Taylor & Francis.
5. Brinkley, P.A. (1994). Characterization of human visual search efficiency for no-decision targets in electronics manufacturing.
6. Carrasco, M., Pizarro, L., & Mery, D. (2010). Visual inspection of glass bottlenecks by multiple-view analysis. *International Journal of Computer Integrated Manufacturing, 23,* 925-941.
7. Carter, C.W. (1957). *Quality control of visual characteristics.* ASQC Convention Transaction, 623-634.
8. Chaney, F.B., & Teel, K.S. (1967). Improving inspector performance through training and visual aids. *Journal of Applied Psychology, 51,* 311-315.
9. Courtney, A.J., & Guan, L. (1998). Assessing search performance with a simulation model. *Human Factors and Ergonomics in Manufacturing, 8,* 251-263.
10. Craig, A. (1983). Vigilance and inspection. *Studia Psychologica, 25,* 259-270.
11. Craig, A. (1984). Human engineering: The control of vigilance. In J.S. Warm (Ed.), *Sustained attention in human performance* (pp. 247-291). Chichester, UK: Wiley.
12. Czaja, S.J., & Drury, C.G. (1981). Training programs for inspection. *Human Factors, 23,* 473-483.
13. Dalton, J., & Drury, C.G. (2004). Inspectors' performance and understanding in sheet steel inspection. *Occupational Ergonomics, 4,* 51-65.
14. Davies. D.R., & Parasuraman. R. (1932). *The psychology of vigilance.* London: Academic Press.
15. Dember, W.N., & Warm, J.S. (1979). *Psychology of perception* (2nd ed.). New York: Holt, Rinehart. & Winston.
16. Drury, C.G. (1974). The human factor in industrial inspection. *Quality Progress VII, 12,* 14-19.
17. Drury, C.G. (1978). Integrating human factors models into statistical quality control. *Human Factors, 20,* 561-572.
18. Drury, C.G. (1989). The information environment in inspection. In *Proceedings of the Second International Conference on Human Factors in Aging Aircraft.* Falls Church, VA: Biotechnology, Inc.
19. Drury, C.G. (1990). Visual search in industrial inspection. In D. Brogan (Ed.), *Visual Search* (pp. 263-276). London: Taylor & Francis.
20. Drury, C.G. (1992). Inspection performance. In G. Salvendy (Ed.), *Handbook of Industrial Engineering (2nd ed.)* (pp. 2282-2314). New York: John Wiley & Sons.

21. Drury, C.G. (1999, December). *Human reliability in civil aircraft inspection.* Paper presented at the RTO HFM Workshop on "The Human Factor in System Reliability—Is Human Performance Predictable?, Siena, Italy.
22. Drury, C.G., & Addison, J.L. (1973). An industrial study of the effects of feedback and fault density on inspection performance. *Ergonomics, 16,* 159-169.
23. Drury, C.G., & Chi, C.F. (1995). A test of economic models of stopping policy in visual search. *IEEE Transactions, 27,* 382-393.
24. Drury, C.G., & Clement, M.R. (1978). The effect of area, density, and number of background characters on visual search. *Human Factors, 20,* 597-602.
25. Drury, C.G., & Fox, J.G. (1975). The imperfect inspector. In C.G. Drury and J.G. Fox (Eds.), *Human Reliability in Quality Control* (pp.11-16). London: Taylor & Francis.
26. Drury, C.G., Karwan, M.H., & Vanderwarker, D.R. (1986). The two-inspector problem. *IIE Transactions, 18,* 174-181.
27. Drury, C.G., & Lock, M.W.B. (1992). Ergonomics in civil aircraft inspection. In E.J. Lovesey, *Contemporary Ergonomics* (pp. 116-123). London: Taylor & Francis.
28. Drury, C.G., & Prabhu, P.V. (1992). Human factors in test and inspection. In G. Salvendy & W. Karwowski (Eds.), *Handbook of Human Factors in Advanced Manufacturing*. New York: Wiley.
29. Drury, C.G., & Prabhu, P. (1996). Information requirements of aircraft inspection: Framework and analysis. *International Journal of Human-Computer Studies, 45,* 679-695.
30. Drury, C.G., & Sinclair, M.A. (1983). Human and machine performance in an inspection task. *Human Factors, 25,* 391-399.
31. Drury, C.G., Spencer, F.W., & Schurman, D.L. (1997). Measuring human detection performance in aircraft visual inspection. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting, 41,* 304-308.
32. Drury, C.G., & Wang, M.J. (1986). Are research results in inspection task specific? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 30,* 476-480.
33. Drury, C.G., & Watson, J. (2002). Good practices in visual inspection. *Aircraft Maintenance Technology*, 74-76.
34. Embrey, D.E. (1979). Approaches to training for industrial inspection. *Applied Ergonomics, 10,* 139-144.
35. Fleck, M.S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. *Psychological Science, 18,* 943-947.
36. Foot, H.C., & Russon, J.M. (1975). The reciprocation of unfavorable evaluations of performance in a two-person inspection task. *European Journal of Social Psychology, 5,* 289-296.
37. Fox, J.G. (1973). Recent human factors contribution to enhancing industrial quality control. *Behaviorometric, 3,* 99-118.
38. Fox, J.G. (1977). Quality control of coins. In H.G. Maule and J.S.Weiner (Eds.). *Case Studies in Ergonomics Practice, Vol. 1*. London, Taylor & Francis.
39. Fox, J.G., & Haslegrave, C.M. (1969). Industrial inspection efficiency and the probability of a defect occurring. *Ergonomics, 12,* 713-721.

40. Gale, A.G., Purdy, K., & Wooding, D.S. (2005). Designing out terrorism: Human factors issues in airport baggage inspection. In D. de Waard, K.A. Brookhuis, R. van Egmond, & Th. Boersema (Eds.), *Human Factors Design, Safety, and Management* (pp. 63-66). Maastricht: Shaker. http://hdl.handle.net/2134/2245
41. Gallwey, T.J. (1982). Selection tests for visual inspection on a multiple fault type task. *Ergonomics, 25,* 1077-1092.
42. Gallwey, T.J. (1998). Evaluation and control of industrial inspection: Part I – Guidelines for the practitioner. *International Journal of Industrial Ergonomics, 22,* 37-49.
43. Gallwey, T.J., & Drury, C.G. (1986). Task complexity in visual inspection. *Human Factors*, *28,* 595-606.
44. Geyer, L.H., Patel, S., & Perry, R.F. (1979). Detectability of multiple flaws. *Human Factors, 21*, 7-12.
45. Ghylin, K.M., Drury, C.G., Batta, R., & Lin, L. (2007). Temporal effects in a security inspection task: Breakdown of performance components. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 51,* 83-97.
46. Ghylin, K.M., Drury, C.G., & Schwaninger, A. (2006, July). *Two-component model of security inspection: Application and findings.* Paper presented at the 16th World Congress of Ergonomics, IEA 2006, Maastricht, The Netherlands.
47. Gonzalez, C., & Madhavan, P. (2011). Diversity during training enhances detection of novel stimuli. *Journal of Cognitive Psychology, 23,* 342-350.
48. Gramopadhye, A., Bhagwat, S., Kimbler, D., & Greenstein, J. (1998). The use of advanced technology for visual inspection training. *Applied Ergonomics, 29*, 361-375.
49. Gramopadhye, A.K., Desai, R.R., Bowling, S., & Khasawneh, M. (2003). Task analysis of general aviation inspection activities: Methodology and Findings. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting, 47,* 36-40.
50. Gramopadhye, A.K., Drury, C.G., & Sharit, J. (1997). Feedback strategies for visual search in airframe structural inspection. *International Journal of Industrial Ergonomics, 19,* 333-344.
51. Gramopadhye, A.K., & Madhani, K. (2001). Visual search and visual lobe size. In C. Arcelli et al. (Eds.), *IWVF4, LNCS 2059* (pp. 525-531). Berlin: Springer-Verlag.
52. Gramopadhye, A.K., Melloy, B.J., Chen, B.J., & Bingham, J. (2000). Use of computer based training for aircraft inspection: Findings and recommendations. *Proceedings of the HFES/IEA Annual Meeting*, San Diego, CA.
53. Gramopadhye, A.K., Melloy, B.J., & Nickles, G.M. (2000). Use of computer based training for aircraft inspection: Minimizing errors and standardizing the inspection process. https://hfskyway.faa.gov/
54. Graybeal, B.A., Phares, B.M., Rolander, D.D., Moore, M., & Washer, G. (2002). Visual inspection of highway bridges. *Journal of Nondestructive Evaluation, 21,* 67-83.
55. Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.
56. Hancock, P.A. (1984). Environmental stressors. In J.S. Warm (Ed.), *Sustained Attention in Human Factors* (pp. 103-142). Chichester: Wiley.

57. Harris, D.H. (1968). Effect of defect rate on inspection accuracy. *Journal of Applied Psychology, 52,* 377-379.
58. Harris, D.H. (1969). The Nature of Industrial Inspection. *Human Factors, 11,* 139-148.
59. Hayashi, E., & Ogawara, Y. (1977). A study on the workload of shear-line inspectors based on eye movement analysis. *Journal of Human Ergology, 6,* 121-126.
60. Hayes, A.S. (1950). Control of visual inspection. *Industrial Quality Control, 6,* 73-76.
61. Heida, J.H. (1989). Characterization of inspection performance. In J. Boogard and G.M. van Dijk (Eds.), *Non-Destructive Testing (Proc. 12th World Conference)* (pp. 1711-1716). Amsterdam: Elsevier.
62. Heidl, W., Thumfart, S., Eitzinger, C., Lughofer, E., & Klement, E.P. (2010, October). *Classifier-based analysis of visual inspection: Gender differences in decision-making.* Paper presented at the 2010 IEEE International Conference on Systems, Man, and Cybernetics, Istanbul, Turkey.
63. Hillman, B.J., Swensson, R.G., Hessel, S.J., Gerson, D.E., & Herman, P.G. (1976). The value of consultation among radiologists. *American Journal of Roentgenology, 127,* 807-809.
64. Jacob, R.J., Raina, S., Regunath, S., Subramanian, R., & Gramopadhye, A.K. (2004). Improving inspector's performance and reducing errors: General aviation training inspection systems (GAITS). *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, 48,* 203-207.
65. Jacobson, H.J. (1952). A study of inspector accuracy. *Industrial Quality Control, 9,* 16-25.
66. Jamieson, G.H. (1966). Inspection in the telecommunications industry: A field study of age and other variables. *Ergonomics, 9,* 297-303.
67. Jerison. H.J. (1963). On the decrement function in human vigilance. In D. N. Buckner & J. J. McGrath (Eds.). *Vigilance: A symposium* (pp. 199-212). New York: McGraw-Hill.
68. Jiang, X., Gramopadhye, A.K., Melloy, B.J., & Grimes, L. (2007). Correlating trust and performance in a hybrid inspection environment.
69. Johnson, S.L., & Funke, D.J. (1980). An analysis of human reliability measures in visual inspection. *Journal of Quality Technology, 12,* 71-74.
70. Juran, J.M., & Gryna, F.M. (1988). Inspection and test. In J.M. Juran (Ed.), *Juran's Quality Control Handbook (3rd ed.).* New York: McGraw-Hill.
71. Kane, J., Moore, D., & Ghanbartehrani, S. (2009). The effect of expectations on visual inspection performance.
72. Kleiner, B.M., & Drury, C.G. (1993). Design and evaluation of an inspection training program. *Applied Ergonomics, 24,* 75-82.
73. Koller, S.M., Drury, C.G., & Schwaninger, A. (2009). Change of search time and non-search time in X-ray baggage screening due to training. *Ergonomics*, *52*, 644-656.
74. Kommidi, S., Dharwada, P., Gramopadhye, A.K., Cho, B.R., & Grimes, L. (2005). Evaluation of human performance in a supervisory inspection task monitoring multiple hybrid inspection systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49,* 573-577.
75. Kraemer, S., Carayon, P., & Sanquist, T.F. (2009). Human and organizational factors in security screening and inspection systems: Conceptual framework and key research needs. *Cognition, Technology & Work*, *11*, 29-41.

76. Kuhlenkotter, B., & Sdahl, M. (2007). Automated inspection system for headlamp reflectors. *The International Journal of Advanced Manufacturing Technology, 32,* 500-504.
77. Latorella, K.A., Gramopadhye, A.K., Prabhu, P.V., Drury, C.G., Smith, M.A., & Shanahan, D.E. (1992). Computer-simulated aircraft inspection tasks for off-line experimentation. *Proceedings of the Human Factors Society 36th Annual Meeting, 36,* 92-96.
78. Lawshe, C.H., & Tiffin, J. (1945). The accuracy of precision instrument measurement in industrial inspection. *Journal of Applied Psychology, 29,* 413-419.
79. Leach, J., & Morris, P.E. (1998). Cognitive factors in the close visual and magnetic particle inspection of welds underwater. *Human Factors, 40,* 187-197.
80. Lion, J.S., Richardson, E., Weightman, D., & Browne, R.C. (1975). The influence of the visual arrangement of material, and of working singly or in pairs, upon performance at simulated industrial inspection. *Ergonomics, 18,* 195-204.
81. Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology, 1,* 6 - 21.
82. Macmillan, N.A., & Creelman, C.D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107,* 401-413.
83. Madhavan, P., Gonzalez, C., & Lacson, F.C. (2007). Differential base rate training influences detection of novel targets in a complex visual inspection task. *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting, 51,* 392-396.
84. McCallum, M., Bittner, A., Rubinstein, J., Brown, J., Richman, J., & Taylor, R. (2005). Factors contributing to airport screener expertise. *Proceeding of the Human Factors and Ergonomics Society Annual Meeting, 49,* 922-926.
85. McCornack, R.L. (1961). Inspector accuracy: A study of the literature. *SCTM 53-61 (14).*
86. McKenzie, R.M. (1958). On the accuracy of inspectors. *Ergonomics, 1,* 258-272.
87. Megaw, E.D. (1979a). The ergonomics of visual inspection. *Applied Ergonomics, 10,* 16.
88. Megaw, E.D. (1979b). Factors affecting visual inspection accuracy. *Applied Ergonomics, 10,* 27-32.
89. Megaw, E.D., & Richardson, J. (1979). Eye movements and industrial inspection. *Applied Ergonomics, 10,* 145-154.
90. Mehta, P., Sadasivan, S., Greenstein, J.S., Gramopadhye, A.K., & Duchowski, A.T. (2005). Evaluating different display techniques for communicating search strategy training in a collaborative virtual aircraft inspection environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49,* 2244-2248.
91. Murgatroyd, R.A., Worrall, G.M., & Waites, C. (1994). *A study of the human factors influencing the reliability of aircraft inspection*. Warrington, UK: AEA Technology.
92. Nuechterlein, K.H., Parasuraman, R., & Jiang, Q. (1983). Visual sustained attention: Image degradation produces rapid sensitivity decrement over time. *Science, 220,* 327-329.
93. Purswell, J.L., & Hoag, L.L. (1974). Strategies for improving visual inspection performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 18,* 397-403.

94. Rao, P., Bowling, S.R., Khasawneh, M.T., Gramopadhye, A.K., & Melloy, B.J. (2006). Impact of training standard complexity on inspection performance. *Human Factors and Ergonomics in Manufacturing, 16*, 109-132.
95. Rasmussen, J. (1983). Skills, rules, knowledge: Signals, signs, and symbols and other distinctions in human performance models. *IEEE Transactions: Systems, Man and Cybernetics, SMC-13,* 257-267.
96. Rebsamen, M., Boucheix, J.M., & Fayol, M. (2010). Quality control in the optical industry: From a work analysis of lens inspection to a training programme, an experimental case study. *Applied Ergonomics, 41*, 150-160.
97. Sadasivan, S., Rele, R., Greenstein, J.S., Gramopadhye, A.K., Masters, J., & Duchowski, A.T. (2005). Collaborative virtual environment to simulate on-the-job aircraft inspection training aided by hand pointing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49,* 2216-2220.
98. Sadasivan, S., Vembar, D., Washburn, C., & Gramopadhye, A.K. (2007). Evaluation of interaction devices for projector based virtual reality aircraft inspection training environments. *Lecture Notes in Computer Science, 4563,* 533-542.
99. Sadeghipour, F., Bugmann, A., Herrera, V., & Bonnabry, P. (2007). The reliability of operators when visually inspecting parenteral drugs. *Practice Research & Innovation, 13,* 41-45.
100. Saito, M., & Tanaka, T. (1977). Visual bottle inspection performance in highly paced belt-conveyor systems. *Journal of Human Ergology, 6,* 127-137.
101. Schoonard, J.W., & Gould, J.D. (1973). Field of view and target uncertainty in visual search and inspection. *Human Factors, 15,* 33-42.
102. Schoonard, J.W., Gould, J.D., & Miller, L.A. (1973). Studies of visual inspection. *Ergonomics, 16,* 365-379.
103. See, J.E., Warm, J.S., Dember, W.N., & Howe, S.R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors, 39*, 14-29.
104. Shagam, R.N., Lerner, J., & Shie, R. (1995). Improved portable lighting for visual aircraft inspection. *SPIE, 2455,* 173-181.
105. Sinclair, M.A. (1979). The use of performance measures on individual examiners in inspection schemes. *Applied Ergonomics, 10,* 17-259.
106. Singh Negi, D. (1981). Pacing, defect rate, and task perception in simulated inspection (Master's thesis). Kansas State University, Manhattan, KS.
107. Slaughter, D.C., Crisosto, C.H., Hasey, J.K., & Thompson, J.F. (2006). Comparison of instrumental and manual inspection of clingstone peaches. *Applied Engineering in Agriculture, 22,* 1-7.
108. Smith, R.L., & Lucaccini, L.F. (1969). Vigilance research: Its application to industrial problems. *Human Factors, 11,* 149-156.
109. Snodgrass, J.G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34-50.
110. Spitz, G., & Drury, C.G. (1978). Inspection of sheet materials: Test of model predictions. *Human Factors, 20,* 521-528.

111. Swain, A.D., & Guttmann, H.E. (1983). *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Application.* Technical Report NUREG/CR-1278-F SAND-0200. Albuquerque, NM: Sandia Corporation.
112. Tanner, W.P., Jr., & Swets, J.A. (1954). A decision-making theory of visual detection. *Psychological Bulletin, 61*, 401-409.
113. Taylor, J.C. (1990). Organizational context for aircraft maintenance and inspection. *Proceedings of the Human Factors Society 34th Annual Meeting, 34,* 1176-1180.
114. Teichner .W.H. (1974) . The detection of a simple visual signal as a function of time on watch*. Human Factors, 16,* 339-353.
115. Thackray, R.I. (1992). *Human factors evaluation of the work environment of operators engaged in the inspection and repair of aging aircraft.* Washington, D.C.: Office of Aviation Medicine.
116. Thackray, R.I. (1993). Correlates of individual differences in nondestructive inspection performance: A follow-up study. http://hfskyway.faa.gov
117. Thomas, L.F., & Seaborne, A.E.M. (1961). The socio-technical context of industrial inspection. *Occupational Psychology, 35,* 36-43*.*
118. Tsao, Y.C., & Wang, T.G. (1984). Inspectors' stopping policy after fault detection. *Human Factors*, *26*, 649-657.
119. Vora, J., Nair, S., Gramopadhye, A.K., Duchowski, A.T., Melloy, B.J., & Kanki, B. (2002). Using virtual reality technology for aircraft visual inspection training: Presence and comparison studies. *Applied Ergonomics, 33,* 559-570.
120. Wales, A.W.J., Anderson, C., Jones, K.L., Schwaninger, A., & Horne, J.A. (2009). Evaluating the two-component inspection model in a simplified luggage search task. *Behavior Research Methods, 41,* 937-943.
121. Wang, M.J., & Drury, C.G. (1989). A method of evaluating inspector's performance differences and job requirements. *Applied Ergonomics, 20,* 181-190.
122. Wang, M.J., & Lin, S.C. (1993). Evaluating the performance of soldering points inspection. In R. Nielsen & K. Jorgensen (Eds.), *Advances in Industrial Ergonomics and Safety V* (pp. 607-614). -11. London: Taylor & Francis.
123. Wang, M.J., Lin, S.C., & Drury, C.G. (1997). Training for strategy in visual search. *International Journal of Industrial Ergonomics*, *20,* 101-108.
124. Warm. J. S. (Ed.). (1984) .An introduction to vigilance .In J. S. Warm (Ed.). *Sustaincd attention in human performance* (pp. l- l4). Chichester: Wiley.
125. Watanapa, A., Kaewkuekool, S., Suksakulchai, S. (2012). Influence of training with and without reward on visual inspector's performance in 3 dimension model. *Applied Mechanics and Materials, 110-116,* 2911-2917.
126. Watts, K.P. (2011). The effect of visual search strategy and overlays on visual inspection of castings (Master's thesis). Iowa State University, Ames, IA. http://lib.dr.iastate.edu/etd/10354
127. Wiener, E.L. (1975). Individual and group differences in inspection. In C.G. Drury and J.G. Fox (Eds.), *Human Reliability in Quality Control* (pp.101-122). London: Taylor & Francis.
128. Wiener, E.I. (1984). Vigilance and inspection. In J.S. Warm (Ed.), *Sustained Attention in Human Performance* (pp. 207-246). Chichester: Wiley.
129. Wilhelmsen, C.A., Ostrom, L.T., & Kanki, B. (2002, August). Aviation visual crack measurement. Presented at *System Safety Society Meeting.* Denver, CO.

130. Woodcock, K. (2003). Task challenges in the performance of amusement ride inspection. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*, *47*, 980-984.
131. Yeow, P.H.P., & Sen, R.N. (2004). Ergonomics improvements of the visual inspection process in a printed circuit assembly factory. *International Journal of Occupational Safety and Ergonomics*, *10*, 369-385.

APPENDIX A: MEASURES OF INSPECTION PERFORMANCE

Table A-1. Empirical Measures of Inspection Performance

Measure	Formula	Description
% Correct Inspections	$A_1 = \frac{H + CA}{Total}$	• Emphasizes how well inspector performs • Treats all responses as equally important • Ignores false alarms and misses
% Satisfactory Product Accepted	$A_2 = \frac{CA}{Total\ Good}$	• Useful when item is costly to produce • 100% performance could occur by accepting everything • Total number of defects does not have to be known
Hit Rate	$A_3 = \frac{H}{Total\ Defects}$	• Measures what inspectors normally see as their job • Useful if FA is unimportant (e.g., rework not a major cost) • 100% performance could occur by rejecting everything • Total number of defects must be known
Efficiency in Improving Product	$E = \frac{\frac{CA}{Total\ Accepts} - \frac{Total\ Good}{Total}}{1 - \frac{Total\ Good}{Total}}$	• Views efficiency as amount by which batch quality improves from inspection • Negative values indicate inspection introduces more errors than it detects • Misses weighted more important than FA • All defects must be known beforehand
Utility-Based Measure	$U = (Cost * M) + (Reward * CA) + (Reward * H) + (Cost * FA)$	• *U* can be made to reflect customer requirements • All response outcomes are taken into account • All defects must be known
Bayesian Measures • Accuracy of Acceptances B_1 • Accuracy of Rejections B_2 • Inaccuracy of Rejections $1 - B_2$	$B_1 = \frac{CA}{Total\ Accepts}$ $B_2 = \frac{H}{Total\ Rejects}$ $(1 - B_2) = \frac{FA}{Total\ Rejects}$	• *A priori* knowledge of defects is not required • B_1 is less useful than E because it does not take into account inspector-induced lowering of quality • B_2 represents what most inspectors regard as an important part of their task, but may lead to acceptance of borderline products

Table A-2. Theoretical Measures of Inspection Performance

Measure	Formula	Description
TSD Perceptual Sensitivity	$d' = z_{FA} - z_H$	• Measures ability to discriminate rejectable items from acceptable items • Normality and variance assumptions must be met[1] • Measures differences between inspectors very well • Reflect influences of most variables in work situations that affect inspection performance (e.g., lighting) • All defects in batch must be known to compute d'
TSD Response Bias	$\beta = \frac{ordinate\ signal-plus-noise\ at\ criterion}{ordinate\ noise\ at\ criterion}$ $c = .5(z_{FA} + z_H)$	• Measures inspector's willingness to respond "reject" • Normality and variance assumptions must be met[1] • All defects in batch must be known to compute β or c

[1] Nonparametric measures of sensitivity and bias are available if the assumptions of normality and variance have been violated or cannot be assessed.

Table A-3. Speed Measures of Inspection Performance

Measure	Formula	Description
Inspection Rate	$R_1 = \frac{\#\ items\ inspected}{Time\ taken}$	• Ignores quality • Encourages inspectors not to inspect to save time
Standardized Inspection Rate	$R_2 = \frac{\#\ items\ inspected * standard\ minutes\ per\ item}{Time\ actually\ taken}$	• Ignores quality • Permits comparison of actual and planned inspection time
Speed/Accuracy Measure	$R_3 = 2(A_1 * R_2)$	• Provides a single index combining both speed and accuracy • High scores could be obtained by very rapid cursory inspection • Low scores may indicate very high quality but slow inspection
Bertelson's Measure	$P = \frac{Time\ available\ for\ inspection}{(\#\ items\ inspected) - (M + FA)}$	• Units are minutes per correct inspection • Low scores indicate good performance • All defects must be known

Table A-4. Classification Measures of Inspection Performance

Measure	Formula	Description
Confusion Matrix	None (see example matrix in Table A-5)	• Matrix readily portrays detected defects, easily missed defects, and classification biases • Main descending diagonal shows correct decisions (highlighted in green in Table A-5) • Off-diagonal cells indicate errors • Errors in the first row of the matrix indicate misses (highlighted in red in Table A-5) • Errors in the first column of the matrix are false alarms (highlighted in yellow in Table A-5)

Table A-5. Example Confusion Matrix

Inspector Response	Actual Batch Composition				
	Good Product	Defect A	Defect B	Defect C	Total
Accept – Good Product	**4700**	**5**	**0**	**45**	4750
Reject – Defect A	**20**	**35**			55
Reject – Defect B	**0**		**20**		20
Reject – Defect C	**80**	10	30	**55**	175
Total	4800	50	50	100	5000

Green = Correct Decisions, Yellow = False Alarms, Red = Misses

This Page Intentionally Left Blank

DISTRIBUTION

1 National Nuclear Security Administration’s Kansas City Plant operated by Honeywell
Federal Manufacturing & Technologies, LLC
Attn: David Boyle
P.O. Box 419159
Kansas City, MO 64141-6159

1 National Nuclear Security Administration’s Kansas City Plant operated by Honeywell
Federal Manufacturing & Technologies, LLC
Attn: John Kendrick
P.O. Box 419159
Kansas City, MO 64141-6159

1 National Nuclear Security Administration’s Kansas City Plant operated by Honeywell
Federal Manufacturing & Technologies, LLC
Attn: David Martin
P.O. Box 419159
Kansas City, MO 64141-6159

1 National Nuclear Security Administration’s Kansas City Plant operated by Honeywell
Federal Manufacturing & Technologies, LLC
Attn: Jan Thompson
P.O. Box 419159
Kansas City, MO 64141-6159

1 National Nuclear Security Administration’s Kansas City Plant operated by Honeywell
Federal Manufacturing & Technologies, LLC
Attn: Rick VanHoose
P.O. Box 419159
Kansas City, MO 64141-6159

1 National Nuclear Security Administration’s Kansas City Plant operated by Honeywell
Federal Manufacturing & Technologies, LLC
Attn: Marie Wallis
P.O. Box 419159
Kansas City, MO 64141-6159

3 NNSA
Attn: Angela Chambers (1)
NA-121.1
Pantex Plant
P.O. Box 30020
Amarillo, TX 79120

1	MS0343	Anna Schauer	2610
1	MS0350	Marcus Craig	2613
1	MS0405	Jared McLaughlin	0433
1	MS0428	Rick Fellerhoff	0400
1	MS0453	Pat Sena	2110
1	MS0481	John Bowers	2116
1	MS0482	Marcey Hoover	0420
1	MS0482	Tommy Woodall	0430
1	MS0483	Nathan Brannon	2112
1	MS0632	John Whitley	2916
1	MS0632	Larry Schoof	2916
3	MS0830	Caren Wenner	0431
1	MS0830	Judi See	0431
1	MS0899	Technical Library	9536 (electronic copy)

Sandia National Laboratories