Training methods for visual inspection tasks

by

Alexander J. Clemons

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Industrial Engineering

Program of Study Committee:

Richard T. Stone, Co-Major Professor

Frank Peters, Co-Major Professor

Scott Chumbley

Iowa State University

Ames, Iowa

2013

TABLE OF CONTENTS

List of Figures....iii
List of Tables....iv
Acknowledgements....v
Abstract....vi
Chapter 1: Introduction....1
1.1 Objective....1
1.2 Search strategies and patterns....2
1.3 Search strategy enhancement....3
1.3.1 Feedback systems....4
1.3.2 Feedforward systems....4
1.3.3 Training of the eye....6
1.4 Cognitive style....7
1.5 Hypothesis....9
Chapter 2: Method....11
2.1 Experimental conditions....11
2.2 Participants....13
2.3 Experimental design....13
2.3.1 Independent variables....13
2.3.2 Dependent variables....13
2.3.3 Additional information....14
2.4 Laboratory setup....15
2.5 Training....16
2.6 Inspection....19
Chapter 3: Results....21
3.1 Signal detection results....21
3.1.1 d' results....22
3.1.2 Hit rate results....23
3.1.3 Correct rejection / false alarm results....25
3.1.4 Repeatability results....27
Chapter 4: Discussion....29
Chapter 5: Conclusion....33
Chapter 6: Future Research....35
References....36

LIST OF FIGURES

Figure 1: MFFT program sample test screenshot 8
Figure 2: Group 2 visual strategy pattern 11
Figure 3: Group 3 visual strategy pattern 12
Figure 4: Eye tracker........ 15
Figure 5: 2XX casting type........ 19
Figure 6: 3XX casting type........ 19
Figure 7: 4XX casting type........ 19
Figure 8: Average individual participant d’ by group; considering all casting types....... 22

LIST OF TABLES

Table 1: Average overall d' result by group....22
Table 2: Two-tailed p-values of group comparisons for overall d' results....23
Table 3: Average d' results by MFFT classification....23
Table 4: Average overall hit-rate, all casting types, by group....24
Table 5: Two-tailed p-values of group comparisons for overall avg. hit-rate, all casting types....24
Table 6: Average overall hit-rate, 4XX casting type, by group....25
Table 7: Two-tailed p-values of group comparisons for overall avg. hit-rate, 4XX casting type....25
Table 8: Average correct rejection rate, all castings types, by group....26
Table 9: Two-tailed p-values of group comparisons for correct rejection rate, all casting types....26
Table 10: Average correct rejection rate, 4XX casting type, by group....26
Table 11: Two-tailed p-values of group comparisons for correct rejection rate, 4XX casting type....26
Table 12: Average standard deviations for hit-rate repeatability, 4XX casting type, by group....27
Table 13: Two-tailed p-values of group comparisons for hit-rate repeatability, 4XX casting type....27
Table 14: Average std. dev. for correct rejection rate repeatability, the 4XX casting type, by group....28
Table 15: Two-tailed p-values of group comparisons for correct rejection rate std. dev. means, 4XX casting type....28

ACKNOWLEDGEMENTS

I would like to thank my major professors, Dr. Richard Stone and Dr. Frank Peters, for their continued support throughout my education at Iowa State University. Their guidance is greatly appreciated. I would also like to thank Dr. Scott Chumbley for his assistance.

I would like to thank Kristopher Watts and many other students at Iowa State University, as well as the Steel Founders Society of America, who all made this study possible.

Finally, I would like to thank my mother and father for their unending encouragement and support.

ABSTRACT

The visual inspection process in the casting industry is a subjective task. Inspectors are often required to make absolute judgments when deciding to classify a casting as acceptable or in need of re-work. Due to this subjectivity, castings can be incorrectly classified as acceptable or scheduled for unnecessary re-work, leading to longer production times or higher costs. It is of interest for both manufacturers and customers of the casting industry to improve the visual inspection process. This research was performed with the goal of using an inspection strategy based on rastering to improve the success of a visual inspection task. Paper and video training methods were used to present a rastering strategy to each participant. The results indicate that video assisted rastering training produced superior results when compared to basic rastering training.

CHAPTER 1: INTRODUCTION

1.1 Objective

Many companies require an inspection of their products at various stages of the manufacturing process. This is done to ensure an acceptable product is produced for the customer. A majority of these inspections are performed by human employees, using only their eyes, often with minimal to no aid from assistive technologies. The manufacturing industry is interested in making the inspection process better.

According to study performed by Daricilar and Peters (2005), when given the task of inspecting castings for defects, the operators studied were able to repeatably identify defects between only 55% and 72% of the time. This can lead to potential unnecessary re-work being performed or unacceptable product being delivered to customers. In order to improve this situation, it must be discovered what is required to have a successful outcome in the visual inspection process. Breaking down the elements of the inspection process gives three basic elements: 1. Knowing what to look for, 2. Knowing how to look for something, and 3. Making a decision. Drury (1990) summarized this by concluding that a successful search requires a decision to be made. This study aimed to explore item 2 above; using a particular search strategy to achieve a desired outcome. The goal was to determine which type of search strategy and training combination resulted in the best visual inspection outcome.

1.2 Search strategies and patterns

A visual inspector must develop a strategy by which he or she will inspect a casting. This strategy can be either random or planned. Planned search strategies are commonly referred to as systematic. Nickles (1998) reveals that a systematic search strategy is the better of the two options in his cursor tracking study. Systematic search strategies can be improved by training a user on how to use a particular strategy.

So that a particular search strategy can be identified as being superior, a method of evaluating the effects of a search strategy must be utilized. To determine the effect of a search strategy that focused on eye movement, Nickles et al. (1998) instructed his participants to use only their eyes to follow a cursor across a predetermined path. The participants were not allowed to use any other guiding aid to assist in following the cursor with their eyes, such as a finger. This allowed for the development of a process to evaluate the performance of an eye-movement-based training strategy. Nickles et al. (1998) discovered there is an improvement in visual inspection performance when a methodical search strategy is used. Hall et al. (2003) confirms this with a study about visual inspection tasks and the impact of search patterns.

While the search pattern is an important factor of a search strategy, the speed at which the inspection is performed is also essential. McDonald and Gramopadhye (1998) studied the influences of decision making speed as well as the relative value of making the correct or incorrect decision. When the speed at which a search pattern is executed increases, the accuracy of the inspection decreases (Koenig et al., 2002). Drury (1994) describes this phenomenon as a trade-off that occurs when either speed or accuracy increases. This means that speed and accuracy are negatively correlated. The study conducted by Koenig et al. (2002) looked at how the speed of an inspection was impacted by various complexity levels of the inspection task. Participants were asked to scan a field of characters and indicate when a particular character was identified. Results of that study provided suitable speeds of cursor movement, which could be extrapolated to infer speeds of eye movement, at the different conditions examined in the study. The conditions of that study which are directly relevant to this study are the density of the background characters and the background characters themselves. The background characters and character density are similar to the anomalies and casting surfaces examined in this study.

Finally, when a search strategy is executed, there are various directions or orientations which each strategy could be performed. An evaluation of three search strategy directions was conducted by Tetteh et al. (2008). The three directions evaluated were horizontal, vertical, and diagonal. The horizontal strategy was found to be superior to the vertical and diagonal orientations. Task complexity and pacing, how fast a person completes the inspection, were also assessed during the study.

1.3 Search strategy enhancement

Search strategies can be augmented by additional systems to improve the results of a visual inspection task. Information regarding the item being inspected as well as appropriate method to perform the inspection could be provided to the inspector to achieve a better result. Drury and Gramopadhye (1992) investigated various methods in which a visual inspection task could be improved. These methods included visual lobe training, feedback systems, feedforward (cueing) systems, as well as the effects of grouping, schema, and attributes. They found that feedback and feedforward systems were the most impactful for improving visual inspection.

1.3.1 Feedback systems

Feedback is a type of information presented to a user or system after an action has been completed. In the realm of visual inspection, it could be information provided to an inspector to aid in having better results with future inspections, such as if something was correctly identified or if the inspector was looking in the correct area. According to Gramopadhye et al. (1997), Gramopadhye et al. (1997), and Ma et al. (2002), there are two types of feedback; performance feedback and cognitive feedback. Performance feedback is concerned with tangible data, including how much of a particular item was viewed by an inspector as well as how much time the inspector took to look at the item. Cognitive feedback is related to presenting information in a meaningful display to the inspector to aid in understanding of inspection performance, often through statistics and graphics (Gramopadhye et al., 1997). Gramopadhye et al. (1997) discovered that those who receive feedback have better results than those who receive no feedback.

1.3.2 Feedforward systems

A feedforward system aims to provide a user with information that is beneficial to acting on a decision, before the action is made. A feedforward system can also be referred to as a cueing system; that is, cueing someone that something is about to happen. This is different than a feedback system as a feedback system relies on the action to have already been made. However, a feedback system and feedforward system can work in tandem to provide a user with useful information for future actions using the results from past actions.

Nickles, et al. (2003) considered three types of feedforward training systems in the realm of visual inspection using systematic search strategies. Participants were divided into three groups; one group per system. All three groups were given verbal instructions to use a systematic search strategy to search for a specified target. The first group was given only this information. The second group was given verbal instructions as well as presented with a static graphic overlay which presented the search strategy in a way such that the user could view the item to be inspected as well as the path which they were to take to inspect the item. The third group was presented with a dynamic overlay system in addition to verbal instruction. The dynamic overlay was similar to the static overlay, with the addition of a means to guide the user along the search path at the proper speed. The results showed the use of all three systems had a positive result on visual inspection performance, but no group was statistically significantly different from either of the other two groups. Watts (2012) also investigated the efficacy of using overlays for visual inspection. Watts's study was noticeably different from that of Nickles's, as the overlay in Watts (2012) was projected onto a physical item in a manner such that the overlay was not consistently in the same location on each casting inspected each time. The overlay used in Nickles (2003) was perfectly static and originated from the same computer screen used for the 2-dimensional inspection task. Watts (2012) found that an overlay did not improve an inspector's performance when compared to basic, non-overlay training, and could be distracting to the inspector if not presented in a way such that the overlay remained in the same orientation and location on the part being inspected. The discrepancy of results between Watts (2012) and Nickles (2003) was likely due to the overlay setup used by Watts (2012).

In a study by Nalanagula, et al. (2006), different feedforward systems were investigated. Like the Nickles, et al. (2003) study, static and dynamic displays were used to gauge visual inspection performance. Nalanagula, et al. (2006) also utilized a hybrid feedforward display that used elements from both a static and dynamic system. The static display offered a view of the overall search pattern to be used, shown overlaid on the part to be inspected. The dynamic display again provided an indicator which the inspector was to follow when inspecting. The hybrid display combined elements of both the static and dynamic systems, providing the inspector the path with a moving indicator as well as a modified display of the static path. It was discovered that dynamic and hybrid displays were superior to a static display and could be used by persons new to visual inspection to improve their results.

1.3.3 Training of the eye

While search patterns and feedback/feedforward systems have been shown to improve visual inspection results, there is an additional factor that must be considered. The human eye itself must be capable of following the search patterns and react to the feedback/feedforward systems. Past research has been conducted that focused on how to train the eye to follow a certain path and pattern. Teaching children how to read is an example of this. Rayner (1978, 1985, 1986) has performed research relating to the teaching of children to read in a particular pattern. Research has also been conducted by

Widdel and Kaster (1981), Rayner (1995), Findley (1997), and Theeuwes, et al. (1999) to investigate the effects of systematic training on the outcome of visual inspection. Similar to the other training method studies mentioned, those who received training had more desirable results than those who received no training.

1.4 Cognitive style

At the core of visual inspection in the casting industry is the individual inspector. Personal differences must be recognized in order to create a system of inspection that will result in the best outcome in a consistent manner. Many studies have attempted to classify visual inspectors based on their cognitive style (Gallwey, 1982; Schwabish and Drury, 1984; Drury and Chi, 1995; Gramopadhye, et al. 1997; Chi and Drury, 1998). All else equal, previous research has shown an individual inspector's cognitive style has a large impact of the inspector's ability to visually inspect a part. A series of ten tests were administered by Gallwey (1982) in hopes of discovering a method to determine if a person would be able to perform an inspection task at a certain level of success. Schwabish and Drury (1984) also classified an inspector based on their cognitive style.

The Schwabish and Drury (1984) study used a test called the Matching Familiar Figures Test, also known as MFFT. The MFFT is a test where a user must match one of six picture options to a given picture. A sample of the MFFT is shown in Figure 1.

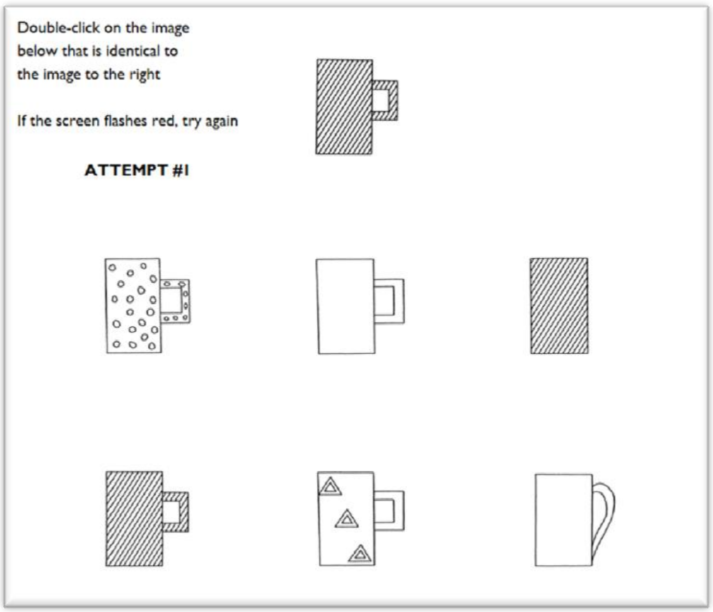

Figure 1: MFFT program sample test screenshot

The MFFT records the amount of time a user takes to make a selection as well as the number of attempts a user makes before selecting the matching figure. Schwabish and Drury (1989) categorized the participants of the study into four groups, established from their MFFT results. Median analysis was used to create the groups, considering the number of attempts to select the correct figure as well as the amount of time taken for each figure. The four groups were classified as fast accurate (FA), impulsive (I), reflective (R), and slow inaccurate (SI). “FA” participants have results that are faster than the median time and less errors than the median error rate. “R” participants have results that are slower than the median time and more accurate than the median error rate. “I” participants have results that are faster than the median and less accurate than the median.

"SA" participants have results that are slower than the median and less accurate than the median.

Wei (2010) performed an MFFT study with a large sample size and found that the FA group was both faster and more accurate in an actual inspection task than the other groups; statistically significantly better than the SI group.

1.5 Hypothesis

Previous literature shows that cognitive style and training methods are both aspects of the visual inspection process that can affect the success of a visual inspection task.

The first hypothesis is the performance of the participants will correlate to their MFFT results, with an outcome similar to what was noted by Wei (2010), that individuals with Fast Accurate MFFT results will perform the inspection task better than those with MFFT results of Slow Inaccurate.

The second hypothesis is that a group of participants given visual inspection training with a video-based training system and a standard left to right, top to bottom search pattern (similar to reading a book) will perform better than a group of participants given basic paper-based training using the same pattern.

The third hypothesis is that a group of participants given visual inspection training with a video training system and a left to right, right to left, top to bottom search pattern (Group 3) will perform worse than the group given video training using a left to right, top to bottom search pattern (Group 2).

A fourth hypothesis is that individuals will have superior inspection performance in regards to repeatability within a particular casting type when given the video-assisted training (Groups 2 and 3) compared to individuals given only verbal training (Group 1).

CHAPTER 2: METHOD

2.1 Experimental conditions

There were three conditions in this experiment. In each condition, the participants were given training on how to properly hold and maneuver a casting, as well as training on what was considered an anomaly that needed to be marked.

The first condition consisted of participants being given castings to inspect with minimal inspection strategy training (Group 1). The second condition consisted of participants that were given a computerized visual search pattern training video consisting of following a cursor on a screen in a left to right, top to bottom fashion, similar to reading a book (Group 2). A screenshot of the PowerPoint slide used for making this video can be seen in Figure 2.

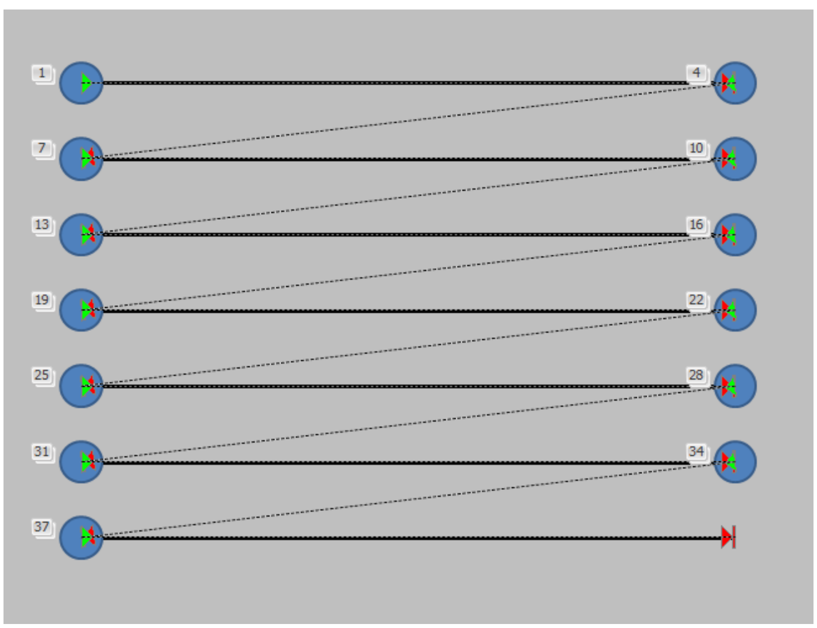

Figure 2: Group 2 visual strategy pattern

The third condition consisted of participants that were given a computerized visual search pattern training video consisting of following a cursor on a screen in a left to right, right to left, top to bottom pattern (Group 3). A screenshot of the PowerPoint slide used for making this video can be seen in Figure 3.

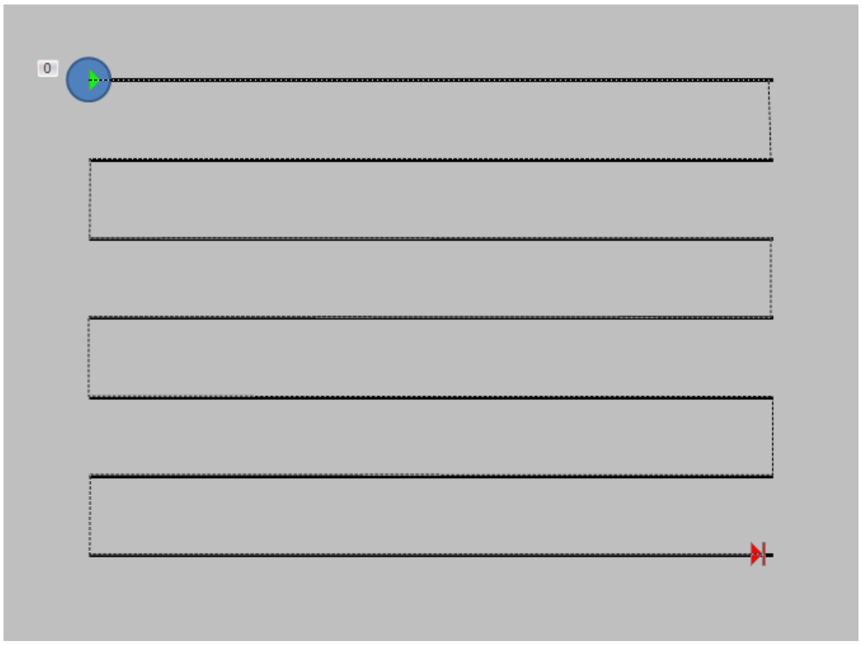

Figure 3: Group 3 visual strategy pattern

Additionally, half of the participants wore eye tracking goggles. The eye tracking goggles were used to check whether or not the participants were using the assigned search strategy. Wearing eye tracking goggles was not a condition of the experiment, as the experimenter determined the goggles did not affect the outcome. Additionally, goggle use was randomly assigned to participants. The participants who did not wear the goggles had their eye movements monitored by the investigator by visually watching the participant's eye movements.

2.2 Participants

There were 12 participants in this study (6 male, 6 female) with an average age of 23.17 (SD = 3.13). All participants were self-selected from the Iowa State University student population and randomly placed into one of three groups by the investigator. Group 1 had an average age of 24 (SD = 2.58). Group 2 had an average age of 23.8 (SD = 4.09). Group 3 had an average age of 21.5 (SD = 1.29). All participants were required to have 20/20 vision, either naturally or corrected.

2.3 Experimental design

2.3.1 Independent variables

There was one independent factor in this experiment: the method of training. There were two methods of search strategy training. One method of search strategy training was the use of paper-based training materials. The second method was using video-based training.

2.3.2 Dependent variables

The dependent variables considered in this experiment are signal detection variables, which include d’, hit rate, and false alarm rate, as well as the cognitive style of the participant. d' is a confounded variable which considers the results of hit rate and false alarm rate. Hit rate is a ratio of the number of correct identifications, or hits, a participant made, of their total attempts, to the ideal number of identifications. A participant could have multiple hits for each item inspected. False alarm rate is the ratio of the number of attempts a participant made at identifying something to the number of those attempts which were incorrect.

2.3.3 Additional information

A metric of calculating signal detection variables was devised using the ASTM A802 SCRATA plates. Levels B4, C1, and J1 were used. The ASTM A802 SCRATA (Steel Castings Research and Trade Association) plates are a set of plastic plates measuring 3" x 5" created to simulate common surface textures of castings. They were developed as a standard to be used for identifying surface features of castings using visual inspection. Level B4 represents surface inclusions, C1 represents gas porosity, and J1 represents welds. So that signal detection theory could be used, each casting to be inspected by the participants was examined using the SCRATA plates as a reference, and the total number of anomalies on each part was recorded in addition to the location of each anomaly.

Signal detection is calculated from the hit rates, false alarm rates, and d'. Sensitivity (d') is calculated as d'= Z(Hit Rate %)-Z(False Alarm Rate) (Stanislaw et al., 1999). According to Stone (2008), d' is "the ratio of the ordinate of the SN (signal plus noise) at the criterion to the ordinate of the random variation of the N (Random variation plus noise) distribution at the criterion."

An MFFT was administered to each participant prior to their training and inspection task. It was administered to each participant using the same computer used for the video training. The screen was a 16:9 15.6" anti-glare, LED lit laptop screen. The MFFT program used was developed at Iowa State University. Each participant was instructed how to properly take the MFFT. Due to the similarity in participant demographics as Watts (2012), the results of the test were calculated using the median analysis found by Watts (2012). These results classified each participant as Slow Inaccurate, Slow Accurate, Fast Inaccurate, or Fast Accurate. In a standard median analysis procedure, any participant with a median score result would be removed from a count. Because no MFFT score in this experiment was equal to the median score found by Watts (2012), no MFFT results were dropped from this experiment.

An eye tracker was constructed using a pair of safety goggles and one Logitech C905 webcam. It was used with six random participants to track their eye motion to ensure the proper search strategy was being used. The eye tracker goggles were treated with an anti-fog solution to prevent they lens from fogging up during inspection. An image of the eye tracker is shown in Figure 4.

Figure 4: Eye tracker

2.4 Laboratory setup

The laboratory for the experiment was set up to allow for the participants to easily inspect the castings at a height comparable to what was found in foundries where actual visual inspection of castings occurs. The experiment was conducted in two laboratories in Black Engineering on the campus of Iowa State University in Ames, IA. One laboratory was used for the administration of the MFFT and the other was used for the inspection of the castings. A series of 49 castings were placed around the perimeter of one lab, grouped by casting type. Each casting was one of three different casting types. The castings used in this study were selected from a catalog casting types provided by foundries to Iowa State University. These castings were selected based on ease of manipulation by the operator, as well as to minimize the likelihood that a pure edge detection strategy would be adopted by the participants while inspecting the castings. A pure edge detection strategy is a visual inspection strategy in which an inspector follows a pattern made of the natural edges of an object while inspecting, not necessarily viewing the entire object. The castings used in this study did not have any physical features or surfaces that would subconsciously direct the participant's eye away from using the search strategy assigned to the participant.

2.5 Training

Search pattern training was given to each of the participants before the inspection task, depending on which group they were assigned to.

Group 1 participants were given visual search pattern training that consisted of instructing the participant to use only their eyes to follow a series of solid and dashed lines printed on a piece of standard copy paper, in a left to right, top to bottom manner, similar to reading a book. First, the participant was given the a piece of copy paper with horizontal solid lines printed the width of the page, with 1" margins on the right and left of the page to allow a uniform background for trailing and preceding movements in edge detection techniques. The lines were spaced approximately 1.3" apart and covered the vertical span of the page, again leaving 1" above the top line and 1" below the bottom line. Each participant was instructed to "read" the horizontal lines for 2 minutes. The participant was then given a different piece of paper to "read" for 2 minutes. This second page featured a layout identical to that of the first page, but the lines were dashed instead of solid.

Group 2 participants were given the same training as Group 1, as well as provided a video to watch. The video featured a cursor moving on the screen in a pattern which with their eyes were to follow when inspecting the castings during the inspection process. The video was provided on an LED backlit computer monitor with a vertical height of 8” and resolution of1366x768. Figure 2 shows the pattern used for Group 2 participants. A series of varying background images were placed behind the cursor pattern. As shown in Figure 2, the initial background was a series of solid lines. These were followed by two series of dashed lines. After the line rastering had concluded, images of the three casting types used in this experiment were underlayed beneath the cursor path. The lines were removed and the cursor followed the same pattern, overlayed on the castings to simulate the pattern to be used by the participant when inspecting each casting type. The cursor completed the pattern three times for each background image and line combination, with the speed of the cursor increasing for each background, each of the three completed paths. The speed of the cursor was reset for each new background. The initial cursor path for each background was completed in 25 seconds (2.03 inches/second). The second time through the pattern, the elapsed completion time was 15 seconds (3.38 inches/second). The final path was completed in 10 seconds (5.08 inches/second). The overall run-time for the video was 5 minutes. Each participant viewed the video twice.

Group 3 participants were shown a video identical to Group 2 participants with one key difference. Rather than the cursor moving from left to right, top to bottom, the cursor in the Group 3 video moved from left to right, then right to left, from the top to the bottom of the screen, as depicted in Figure 3. Again, the video run-time was 5 minutes and each participant viewed the video twice. Group 3 participants also received the same training as Group 1.

Anomaly training was given to each participant immediately prior to their inspection task, and after the search pattern training. An anomaly in this usage was a part of the casting which was considered an irregularity by the investigator, determined using SCRATA identifiers (Section 2.6.3). As part of the anomaly training, the participants inspected 2 castings of each casting type (Figures 5, 6, and 7). During this phase of training, the participants were allowed to ask questions regarding whether or not certain features of the casting were anomalies. The purpose of this was to allow the participants to become familiar with the castings they would be inspecting as a part of their task, as well as the anomalies they were to be looking for.

After this was complete the participants were instructed to begin the inspection task.

2.6 Inspection

Each participant was asked to inspect a total of 81 parts from three different types of castings, noted throughout as 2XX, 3XX, and 4XX. The first digit represented the casting type, followed by 2 digits to identify each casting. The 2XX casting type (Figure 5) was round and bell-shaped, approximately 8” in diameter and 7 pounds in weight. The 3XX casting type (Figure 6) was approximately 10” x 7” x 1”, weighing 3 pounds. The 4XX casting type (Figure 7) was approximately 11” x 8” x 2.5” It weighed approximately 5 pounds.

Figure 5: 2XX casting type

Figure 6: 3XX casting type

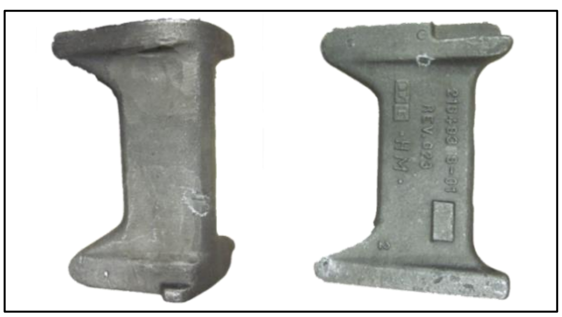

Figure 7: 4XX casting type

Each casting was initially inspected by the investigator to determine the number and location of anomalies on each casting. The SCRATA plates were used for this purpose. Some castings had no anomalies. The castings were then photographed and cataloged into a computer for later comparisons to be made.

The castings were arranged around the perimeter of the lab, grouped by casting type, and placed in a random order within each casting type. The inspector was asked to circle any anomalies found with a piece of white chalk. The circles were to be 1" in diameter. If an anomaly was larger than 1" in diameter, the inspector was asked to draw enough circles to contain the anomaly. White chalk was used because it was easy to remove from the castings when the inspector had finished, and stayed on the casting long enough for the investigator to determine where on the casting the participant had marked an anomaly. Due to the arrangement of the castings, the participants were instructed not to look at any previous castings they had marked, or any future castings they would be marking, when deciding whether or not a particular part had any anomalies that required marking. This was to prevent relative judgments from being made. It should be noted that two participants were provided parts on a conveyance system identical to that of Watts (2012). This system involved providing the castings in groups of 2 to 4 to the inspectors on a wheeled piece of wood, 24" square. The inspectors would then inspect the castings and when finished, a new set of castings would be provided to the inspector. The height of inspection and opportunity for relative judgments were equal with the conveyance system, so the data was used. The change was made from the conveyance system to the perimeter system to ease the data collection process by the investigator. Each casting inspected was photographed by the investigator for later analysis.

CHAPTER 3: RESULTS

3.1 Signal detection results

After the participants had inspected each of the castings, the investigator determined the number of hits and false alarms for each casting in each casting type group. This was done by comparing photographs the investigator captured of each inspected casting with the photographs captured of the castings which were marked by the investigator using the SCRATA plates as a standard. d' was then calculated using the hit rate and false alarm rate.

The unpaired Student's T-Test was used to analyze all results to determine any differences between the three groups. The t value was calculated using the formula shown below:

$$t = \frac{Signal}{Noise} = \frac{\overline{x}_1 - \overline{x}_2}{S_{pooled}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$S_{pooled} = \sqrt{\frac{s_1^2 (n_1 - 1) + s_2^2 (n_2 - 1)}{n_1 + n_2 - 2}}$$

(University of Michigan Engineering Department, 2012)

Where:

$\overline{x}_1$ is the average of the first data set

$\overline{x}_2$ is the average of the second data set

n_1 is the number of measurements in the first data set

n_2 is the number of measurements in the second data set

s_1 is the standard deviation of the first data set

s_2 is the standard deviation of the second data set

This version of the t-test was used because the sample means, standard deviations, and number of participants in each group are known.

Each group was compared against each of the other two groups. Group 1 (paper training) was also compared to the combined results of Groups 2 & 3 (video training).

3.1.1 d' results

Table 1 shows the average overall d' results for each of the three groups, considering the visual inspection results for all three casting types inspected by each participant in each group. A higher d' value represents an overall better inspection score.

Table 1: Average overall d' result by group

	Group 1	Group 2	Group 3	Groups 2 & 3
Average d'	0.534	0.442	0.582	0.512

Figure 8 shows the d' results of each participant in each group. Each group had four participants.

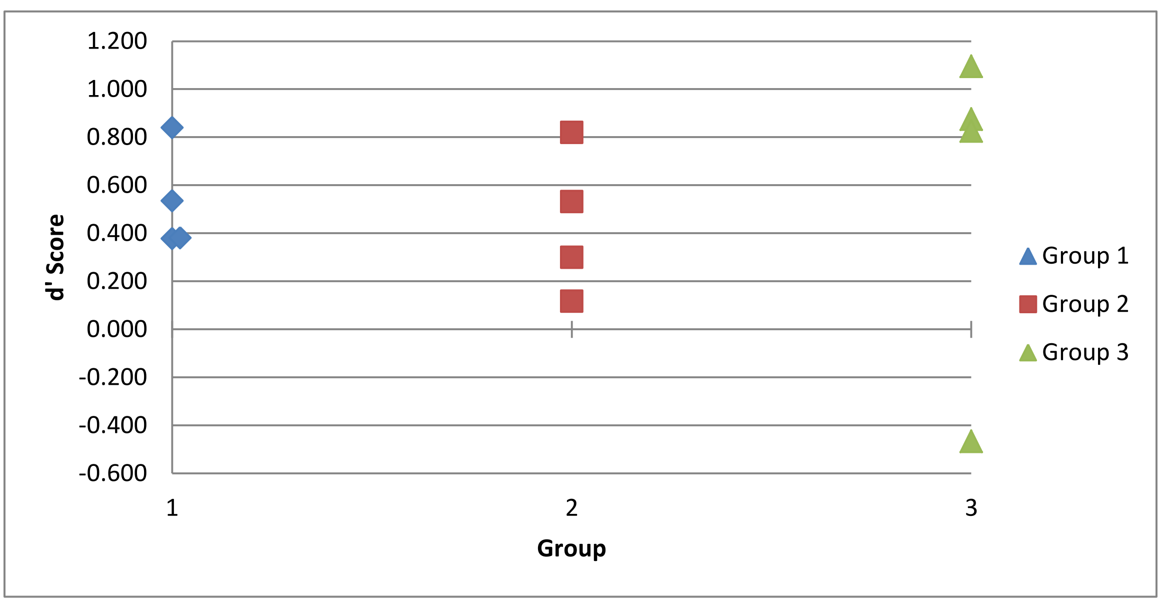

Figure 8: Average individual participant d' by group; considering all casting types

Table 2 lists two-tailed P-values for comparisons between d' results for the different training groups. The unpaired Student's T-Test was used because the sample means, standard deviations, and number of participants in each group are known. Also, the Group d' values used in the calculation of the t statistic were averages of the average d' values for each participant and could be assumed to follow a normal distribution according to the Central Limit Theorem. There was not a significant difference in the d' results between the groups ($\alpha=0.10$). The p-value corresponds to the probability that a test-statistic would be greater than or equal to the values observed. A p-value closer to 0 would correlate to a significant difference in the samples.

Table 2: Two-tailed p-values of group comparisons for overall d' results

	Group 1 vs. 2	Group 1 vs. 3	Group 2 vs. 3	Group 1 vs. 2&3
P-value	0.641	0.902	0.730	0.927

3.1.1.1 d' –MFFT results

Table 3 shows the average d' result for each classification of inspector type, considering all casting types and all parts inspected. There was not a statistical significant difference between the groups.

Table 3: Average d' results by MFFT classification

Group	Slow Inaccurate	Slow Accurate	Fast Inaccurate	Fast Accurate
Number of Participants	1	6	3	2
d' result	0.300	0.522	0.580	0.736

3.1.2 Hit rate results

Hit rate was measured for each participant, considering all casting types, as well as for the 4XX casting type alone. Separate calculations were done for the 4XX casting type, as this was the casting type used for repeatability calculations. Each participant inspected each of the 4XX castings a total of three times.

3.1.2.1 Overall hit rate – all casting types

Table 4 shows the average value of a successfully identified "hit" when considering all casting types. Table 5 shows the P value for the comparison between groups for a successfully identified "hit" when considering all casting types. There was not a significant difference between any of the groups (α=0.10).

Table 4: Average overall hit-rate, all casting types, by group

	Group 1	Group 2	Group 3
Average	0.746	0.813	0.794

Table 5: Two-tailed p-values of group comparisons for overall avg. hit-rate, all casting types

	Group 1 vs. 2	Group 1 vs. 3	Group 2 vs. 3	Group 1 vs. 2&3
P-value	0.462	0.720	0.878	0.496

3.1.2.2 Overall hit rate – 4XX castings

Table 6 shows the average value of a successfully identified "hit" when considering the 4XX casting type. Again, this casting type was inspected by each participant a total of three times. The 4XX casting type was used for repeatability measurements. Table 7 shows the P value for the comparison between groups for a successfully identified "hit" considering the 4XX casting type. While the P-values were lower for the 4XX casting types than when considering all casting types, there was again not a significant difference between any of the groups (α=0.10).

Table 6: Average overall hit-rate, 4XX casting type, by group

	Group 1	Group 2	Group 3
Average	0.652	0.773	0.788

Table 7: Two-tailed p-values of group comparisons for overall avg. hit-rate, 4XX casting type

	Group 1 vs. 2	Group 1 vs. 3	Group 2 vs. 3	Group 1 vs. 2&3
P-value	0.363	0.360	0.903	0.206

3.1.3 Correct rejection / false alarm results

The correct rejection rate and the related measurement of false alarms were measured for each participant considering all casting types, as well as for the 4XX casting type alone. The unpaired Student's T-Test was used to help determine whether the sample means of the correct rejection rate were significantly different between the groups. The standard deviations were approximately equal, and the sample means, standard deviations, and number of participants in each group were known. The correct rejection rate is directly related to the false alarm rate. The false alarm rate can be calculated by subtracting the correct rejection rate from 1.

3.1.3.1 Correct rejection rate – all casting types

Table 8 shows the average value of the correct rejection rate when considering all casting types. Table 9 shows the P value for the comparison between groups for the correct rejection rate when considering all casting types. When comparing the group with standard raster training (Group 1) with the groups that received one of the video training (Groups 2 & 3), there was a significant difference (α=0.10). This means participants who received the video training had a higher likelihood of not marking the casting where it should not have been marked.

Table 8: Average correct rejection rate, all castings types, by group

	Group 1	Group 2	Group 3
Average	0.567	0.680	0.652

Table 9: Two-tailed p-values of group comparisons for correct rejection rate, all casting types

	Group 1 vs. 2	Group 1 vs. 3	Group 2 vs. 3	Group 1 vs. 2&3
P value	0.207	0.257	0.655	**0.094**

3.1.3.2 Correct rejection rate – 4XX castings

Table 10 shows the average value of the correct rejection rate when considering the 4XX casting type. Table 11 shows the P value for the comparison between groups for the correct rejection rate when considering only the 4XX casting type. Similar to the comparison of correct rejections when considering all casting types, the lowest P value in this comparison was between the group with standard raster training and the groups with video training. However, there was not a significant difference ($\alpha=0.10$).

Table 10: Average correct rejection rate, 4XX casting type, by group

	Group 1	Group 2	Group 3
Average	0.676	0.765	0.748

Table 11: Two-tailed p-values of group comparisons for correct rejection rate, 4XX casting type

	Group 1 vs. 2	Group 1 vs. 3	Group 2 vs. 3	Group 1 vs. 2&3
P-value	0.318	0.345	0.773	0.169

3.1.4 Repeatability results

The standard deviations in both hit rate and correct rejection rate were calculated for each participant considering each participant's three trials of inspecting the 4XX castings. The unpaired Student's T-Test was used to help determine whether the sample means of the standard deviations were significantly different between the groups.

3.1.4.1 Hit rate

Table 12 shows the average value of the standard deviations of each group when considering the repeatability of hit rate for the 4XX casting type. Table 13 shows the P value for the comparison between groups for hit rate repeatability when considering the standard deviations of the hit rates for each trial of each participant's inspection of the 4XX casting type. There is a significant difference (α=0.10) between the average standard deviation of the hit rate between the group with basic raster training (Group 1) and the groups with video raster training (Groups 2 & 3). This means that Groups 2 and 3 had a higher likelihood of having a better hit rate than Group 1.

Table 12: Average standard deviations for hit-rate repeatability, 4XX casting type, by group

	Group 1	Group 2	Group 3
Average	0.162	0.062	0.062

Table 13: Two-tailed p-values of group comparisons for hit-rate repeatability, 4XX casting type

	Group 1 vs. 2	Group 1 vs. 3	Group 2 vs. 3	Group 1 vs. 2&3
P-value	0.131	0.153	0.473	**0.033**

3.1.4.2 Correct rejection rate

Table 14 shows the average value of the standard deviations of each group when considering the repeatability of the correct rejection rate for the 4XX casting type. Table 15 shows the P value for the comparison between groups for correct rejection rate repeatability when considering the standard deviations of the correct rejection rates for each trial of each participant's inspection of the 4XX casting type. There is a significant difference (α=0.10) between the average standard deviation of the correct rejection rate between the group with basic raster training (Group 1) and the groups with video training (Groups 2 & 3).

Table 14: Average std. dev. for correct rejection rate repeatability, the 4XX casting type, by group

	Group 1	Group 2	Group 3
Average	0.087	0.044	0.024

Table 15: Two-tailed p-values of group comparisons for correct rejection rate std. dev. means, 4XX casting type

	Group 1 vs. 2	Group 1 vs. 3	Group 2 vs. 3	Group 1 vs. 2&3
P-value	0.214	**0.049**	0.264	**0.026**

CHAPTER 4: DISCUSSION

The MFFT data shows there is a descriptive difference in cognitive style when considering d' results of the inspection process. However, there was only one participant with an MFFT result of SI. As such, the sample size of each MFFT classification made statistical inferences meaningless, and no statistical tests were performed. However, when comparing the d' results associated with each MFFT classification to the results found in Watts (2012), the trend of higher d' scores being associated with Fast Accurate inspectors holds true. The d' values found in the study for MFFT classifications of Slow Accurate and Fast Inaccurate are also roughly equal with the results found in Watts (2012). Considering no statistical significance was found in this study for the d' results for each MFFT classification, the first hypothesis can be neither proved nor disproved. Additionally, the sample of participants used in this study was entirely of college students, which may have a differing MFFT classification distribution than what one might expect to find in the casting industry.

When a visual inspector uses a systematic search strategy, there can be improvement in performance (Nickles, et al., 1998), which, related to this study, can lead to increased successful inspection outcomes. Furthermore, an improvement can also be expected when an individual is trained to use a specific strategy (Findley, 1997; Rayner, 1995; Theeuwes, et al., 1999; Widdel and Kaster, 1981). This experiment confirmed these ideas. As the amount of training increased, the inspector's rate of false alarms decreased and their ability to consistently mark castings increased. It is possible to partially confirm the second hypothesis. Video training as a whole (considering both

Groups 2 and 3) provided statistically significantly better visual inspection outcomes than that of Group 1 (non-video rastering training). However, considering the comparison of visual inspection outcomes of Group 2 to Group 1 and of Group 3 to Group 1, there were not statistically significant results. This means there may or may not have been an improvement in visual inspection outcomes as a result of video rastering training.

When making conclusions about the effectiveness of the different types of training provided to the participants, no statistically significant results were found while considering only d'. d' is a confounded score of many aspects of the inspection process. Therefore it was necessary to look at the individual components that compose d'. The correct rejection rate is a measure of the ability of the inspector to recognize when something does not need to marked. When considering all casting types, the video training method (Groups 2 and 3) provided a statistically significant mean correct rejection rate that was better than that of the standard training group (Group 1). This further confirms the idea that an improvement in performance can be expected when an individual is trained to use a particular strategy (Findley, 1997; Rayner, 1995; Theeuwes, et al., 1999; Widdel and Kaster, 1981). The additional method of training (video of search patterns) provided the participants an additional medium of training, suggesting that active training may produce better results than passive training.

The sample size of each group was four participants. When performing t statistic calculations, this leads to a relatively small degree of freedom. Consequently, it is rather difficult to prove significance at a level usually accepted by conventional criteria. It is worth noting that the p values found in this study, while rarely statistically significant, were generally in favor of suggesting better performance among the video training groups (Groups 2 and 3).

The third hypothesis, stating that Group 3 will perform worse than Group 2, was unable to be proven. Americans are typically taught to read following a left to right, top to bottom pattern. The eyes are familiar with that pattern and the training given to Group 2 builds off of that education. Group 3 was asked to stray from this and adopt a different strategy; following a pattern that moved from left to right, then right to left, top to bottom. As unnatural as this pattern may initially seem, it was employed by many visual inspectors found in the casting industry, mainly due to the large castings sizes they must inspect. This pattern also makes use of eye travel back to the left side of an object. While the castings in this experiment were considerably smaller than what some inspectors examine in the field, the same principles apply. However, the training method provided to Group 3 forgoes typical edge detection. Even though the rastering pattern for Group 3 makes use of the eye movement from right to left, using that movement for inspection can be an unnatural process for those who are used to inspecting from left to right. It was therefore believed by the investigator that participants in Group 3 would perform worse than those in Group 2. The P values of the comparisons of means between Group 2 and Group 3 were usually the highest among all comparisons, concluding that there was likely little difference in the performance of Group 2 and Group 3.

The repeatability of the inspector's performance was measured using the 4XX casting type while considering the hit rate and correct rejection rate standard deviations of each inspection trial set. It was found that Groups 2 and 3 performed more consistently than Group 1 when considering both hit rate and correct rejection rate. The fourth hypothesis was proved to be correct.

CHAPTER 5: CONCLUSION

This study had four objectives. The first objective was to determine if the performance of the participants will correlate to their MFFT results. The second was to determine if a video raster training system would perform better than a group of participants given basic training using the same pattern. A third objective was to discover if visual inspection training with a left to right, right to left, top to bottom search pattern would perform worse than a group given training using a left to right, top to bottom search pattern. The fourth objective was to find if superior inspection performance in regards to repeatability would be realized with video-assisted training compared to individuals given only verbal training.

This study was unable to prove the first objective of correlation between MFFT result and performance statistically true. However, the d' scores were consistent with those found by Watts (2012). It should also be noted that d' scores were higher in the Fast Accurate group than those of the Slow Inaccurate group.

This study was able to determine that there is an increase in visual inspection performance when the participants were given a video raster training system compared to those who were given basic raster training. The correct rejection rate was statistically significantly better in the groups given the video training than those who received the basic training. There was an increase of 7.2% in the ability of an inspector to correctly identify the absence of an anomaly. Using either video training method should provide the casting industry an improved training method to increase on-the-job performance of a visual inspection task.

In respect to the third objective of determining if a raster pattern similar to reading a book would be superior to that of an unfamiliar pattern, there was no significant difference between the two patterns. Considering both patterns were relatively similar and were administered using the video training method, one may conclude that the use of video training has more impact on the outcome of training than on the actual search pattern.

The fourth and final objective was to determine if a participant would be able to increase their repeatability when making a decision of whether or not to mark something on a casting. This study was able to find that a video-based training method increased the repeatability of a participant's decision making. This is important to the casting industry as it could lead to increased effectiveness of remedial training exercises due to the inspector's superior consistency.

CHAPTER 6: FUTURE RESEARCH

Future research related to this study would benefit from a larger sample size of each group. If the sample size of each group in this study were larger, there would have likely been more significance found in the results.

It would also be interesting to perform this research in a foundry setting rather than in a laboratory setting. This may be especially interesting in relation to repeatability measures. Using inspectors from a foundry may also lead to the discovery of a stronger correlation between MFFT results and inspector performance.

An additional aspect that could be researched in the future is the consideration of substantially larger castings. The rastering patterns employed in this study were relatively easy to maintain when considering the casting sizes.

REFERENCES

Blatter, K., Graw, P., Munch, M., Knoblauch, V., Wirz-Justice, A., & Cajochen, C. (2006). Gender and age differences in psychomotor vigilance performance under differential sleep pressure conditions. Behavioral Brain Research, 168, 312– 317.

Chi, C.-F. and Drury, C. G. (1998). Do people choose an optimal response criterion in an inspection task? IIE Transactions, 30, 257-266.

Czaja, S. J. and Drury, C. G. (1981). Training programs for inspection. Human Factors, 23(4), 473-484.

Daricilar, G. and Peters, F. (2005). Measurement Error of Visual Casting Surface Inspection. Proceedings of the 2005 Steel Founders' Society of America – Technical and Operating Conference. Chicago IL.

Drury, C. G. (1990). Visual search in industrial inspection. In D. Brogan (Ed.), Visual search (pp. 263-276). London: Taylor & Francis.

Drury, C. G., and Gramopadhye, A. K. (1992). Training for visual inspection: Controlled studies and field implications. Proceedings of the Seventh Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection, Washington, DC: Office of Aviation Medicine, 135-146.

Drury, C. G. (1994). The speed-accuracy trade-off in industry. Ergonomics, 37, 747-763.

Drury, C. G. and Chi, C.-F. (1995). A test of economic models of stopping policy in visual search. IIE Transactions, 27, 382-393.

Findley, J. M. (1997). Saccade target selection during visual search. Vision Research, 37, 617–631.

Gallwey, T. J. (1982). Selection tests for visual inspection on a multiple fault type task. Ergonomics, 25, 1077-1092.

Gallwey, T. J. and Drury, C. G. (1986). Task complexity in visual inspection. Human Factors, 28(5), 595-606.

Gramopadhye, A. K., Drury, C. G., and Prabhu, P. V. (1997). Training strategies for visual inspection. Human Factors and Ergonomics in Manufacturing, 7(3), 171-196.

Gramopadhye, A. K., Drury, C. G., and Sharit, J. (1997). Feedback strategies for visual search in airframe structural inspection. International Journal of Industrial Ergonomics, 19, 333-344.

Hall, R., Jiang, X., Bowling, S. R., Khasawneh, M. T., Kaewkuekool, S., and Gramopadhye, A. K., "An Evaluation for Job-Aiding for an Inspection Task," Proceedings of the Human Factors and Ergonomics Society, Denver, October 2003.

Kaufman, J., Gramopadhye, A. K., and Kimbler, D. (2000). Using training to improve inspection quality. Quality Engineering, 12(4), 503-518.

Koenig, S., Gramopadhye, A. K., and Melloy, B. J. (2002). Use of a Job Aid to Promote Systematic Search Under Different Levels of Task Complexity. Human Factors and Ergonomics in Manufacturing, 12(4), 349-363.

Ma, J., Drury, C. G., and Bisantz, A. M. (2002). Impact of feedback training in CBT in visual inspection. In Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, Santa Monica, CA, 2079-2083.

McDonald, T. and Gramopadhye, A. K. (1998). The effect of pacing and explicit costs on human decision making performance in inspection. In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Santa Monica, CA, 2079-2083.

Nalanagula, D., Greenstein, J. S., Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. International Journal of Industrial Ergonomics, 36, 289-300.

Nickles, G. M., Sacrez, V., Gramopadhye, A. K. (1998). Can we train humans to be systematic inspectors? In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Santa Monica, CA, 1165-1169.

Nickles, G. M., Melloy, B. J., and Gramopadhye, A. K. (2003). A comparison of three levels of training designed to promote systematic search behavior in visual inspection. International Journal of Industrial Ergonomics, 32, 331-339.

Rayner, K. (1978). Eye movements in reading and information processing. Psychological Bulletin, 85, 618–660.

Rayner, K. (1985). The role of eye movements in learning to read and reading disability. Remedial and Special Education, 6, 53–60.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. Journal of Experimental Child Psychology, 41, 211–236.

Rayner, K. (1995). Eye movements and cognitive processes in reading, visual search, and scene perception. In J. M. Findlay , R. Walker , & R. W. Kentridge (Eds.), Eye movement research: Mechanisms, processes and applications (pp. 3–22). Amsterdam: North Holland.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. The Quarterly Journal of Experimental Psychology, 62:8, 1457-150.

Rousek, J. and Hallbeck, M. (2011). Improving and analyzing signage within a healthcare setting. Applied Ergonomics, 42:6, 771-784.

Sadasivan, S., Greenstein, J. S., Gramopadhye, A. K., and Duchowski, A. T., (2005). Use of eye movements as feedforward training for a synthetic aircraft inspection task, In Proceedings of the ACM CHI 2005 Conference, Portland, OR, 141-149.

Schwabish, S. D. and Drury, C. G. (1984). The influence of the reflective-impulsive cognitive style on visual inspection. Human Factors, 26(6), 641-647.

Stone, R. T. (2008). Augmented multisensory interface design: Performance enhancement capabilities and training potential. PhD dissertation. Amherst, N.Y.: State University of New York at Buffalo.

Tetteh, E., Jiang, X., Mountjoy, D., Seong, Y., McBride, M. (2008). Evaluation of a job-aiding tool in inspection systems. Human Factors and Ergonomics in Manufacturing, 18(1), 30-48.

Theeuwes, J., Kramer, A. F., Hahn, S., Irwin, D. E., & Zelinsky, G. J. (1999). Influence of attentional capture on occulomotor control. Journal of Experimental Psychology: Human Perception and Performance, 25(6), 1595-1608.

Wei C.S., (2010). The impact of inspector’s cognitive style on performance in various visual inspection display tasks. (Master’s thesis).

Widdel, H. and Kaster, J. (1981). Eye movement measurement in the assessment and training of visual performance, Manned systems design: Methods, equipment, and applications; Proceedings of the Conference, Freiburg im Breisgau, West Germany; United States; 22-25 September, 251-270.